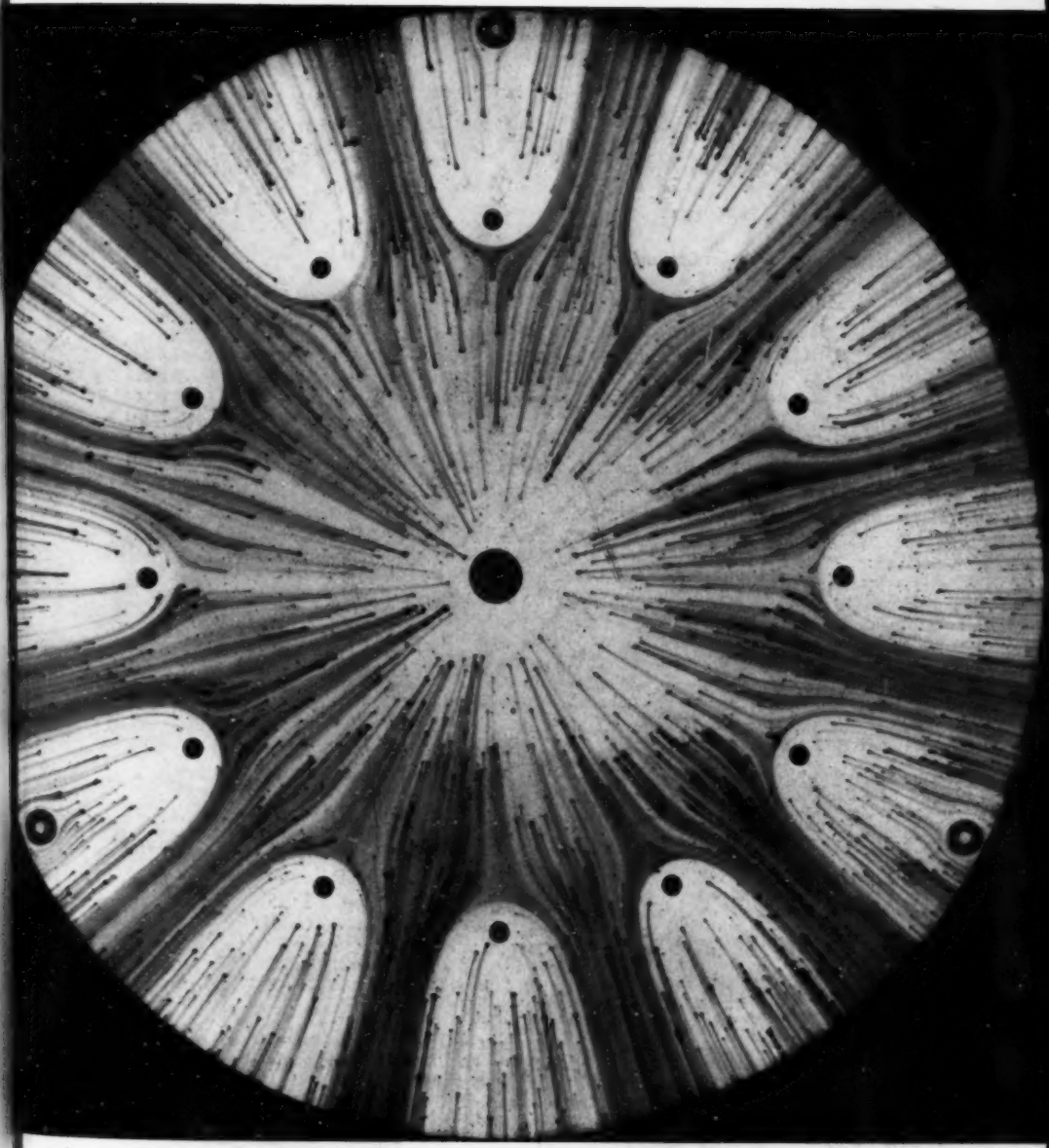
JOURNAL OF APPLIED PHYSICS



Fluid Flow Mapping Used by Professor A. D. Moore to Simulate Fields Occurring in Electrostatic, Magnetic, Electric Current, Heat Flow, Chemical Diffusion, and Other Phenomena. **(See page 790.)**

JOURNAL OF APPLIED PHYSICS

Vol. 20, No. 8 AUGUST, 1949

In This Issue

	Page
Sixth Annual Pittsburgh Conference on X-Ray and Electron Diffraction	725
Precision Determination of Lattice Constants by the Powder and Rotating Crystal Methods and Applications M. E. Straumanis	726
Dead Time and Non-Linearity Characteristics of the Geiger-Counter X-Ray Spectrometer Leroy Alexander, Elizabeth Kummer, and Harold P. Klug	735
Reflection Method of Determining Preferred Orientation on the Geiger-Counter Spectrometer Michael Field and M. Eugene Merchant	741
Influence of Magnetic Field on Recrystallization R. Smoluchowski and R. W. Turner	745
Morphology of Carbon Black Particles in Shadow-Cast Specimens John H. L. Watson	747
Coexistence Relations of $n+1$ Phases in n-Component Systems Paul M. Pepper	754
On the Diffraction of an Electromagnetic Wave through a Plane Screen J. W. Miles	760
Method for the Instantaneous Measurement of Velocity and Temperature in High Speed Air Flow D. G. Marlow, C. R. Nisewanger, and W. M. Cady	771
Some Properties of the Ba_2SiO_4 Oxide Cathode Interface A. Eisenstein	776
Fields from Fluid Flow Mappers A. D. Moore	790
High Inverse Voltage Germanium Rectifiers S. Benzer	804
Letters to the Editor:	
Ordering Reaction in Co-Pt Alloys J. B. Newkirk, A. H. Geisler, and D. L. Martin	816
Impedance Synthesis without Use of Transformers R. Bott and R. J. Duffin	816
Improvement in the Shadow-Cast Replica Technique S. J. Singer and R. F. Petzold	816
Subcooling of Liquid Metals D. Turnbull	817
Measurement of the Dielectric Constant and Loss of Solids and Liquids by a Cavity Perturbation Method George Birnbaum and Jacques Franeau	817
Note on "The Mixing Properties of Non-Linear Condensers" R. V. L. Hartley	819
Note on "Qualitative Experimental Verification of the Change of Burning Rate of Rocket Powders with Radiation Path Length" Henri Muraour	819
Reply to Muraour's Letter to the Editor S. S. Penner	819
Erratum: Photo-Density Method of Determining the Relative Penetration of Diffused Sodium 24 Tracer into Glass James R. Johnson	819
New Books	820
Index to Advertisers	xxxii

BOARD OF EDITORS
Elmer Hutchisson, *Editor*

Associate Editors
R. Bowling Barnes
Otto Beeck
W. James Lyons
Philip Morrison
Paul E. Sabine
Raymond J. Seeger
Frederick Seitz
Charles S. Smith
J. R. Townsend

Publication of the American Institute of Physics

OFFICERS
George R. Harrison, *Chairman*
Wallace Waterfall, *Secretary*
George B. Pegram, *Treasurer*
John T. Tate, *Adviser on Publications*

GOVERNING BOARD
Luis W. Alvarez
J. W. Beams
Wallace R. Brode
E. U. Condon
W. F. Fair, Jr.
Harvey Fletcher
George R. Harrison
W. V. Houston
Paul Kirkpatrick
Paul E. Klopsteg
F. W. Loomis
Philip M. Morse
J. R. Oppenheimer
George B. Pegram
Harold K. Schilling
John T. Tate
Wallace Waterfall
M. W. White
A. G. Worthing

Consultant on Printing
Melvin Loos

ADMINISTRATIVE STAFF
Henry A. Barton, *Director*
Theodore Vorburger, *Business Manager*
Ruth F. Bryans, *Publication Manager*
Betty Hofstadter, *Office Manager*
Eileen B. Neuberger, *Circulation Manager*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. The authors' institutions are requested to pay a **publication charge** of $4 per page which, if honored, entitles them to 100 free reprints. Instructions will be sent with galley proofs.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$7.00	$ 8.00
To all others	9.00	10.00

Back Number Prices
Complete set: Vol. 1, July, 1931–Vol. 19, 1948—$140.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1949, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 20, Number 8 August, 1949

Sixth Annual Pittsburgh Conference on X-Ray and Electron Diffraction

THE sixth of a series of conferences on x-ray and electron diffraction organized each year by a group of Pittsburgh scientists independently of any professional society was held last November. These conferences have grown rapidly in importance and now are the leading regular meetings in their field in this country. This time the occasion was highlighted by a visit of Sir Lawrence Bragg, who delivered the main lecture, on "The X-Ray Structure of Proteins and Other Organic Molecules." It is hardly necessary to describe the beauty and importance of this domain of research and the fundamental character of the results obtained by Sir Lawrence and his collaborators. The interplay of observation, analytical deduction, and sheer intuition in disentangling the vast field of protein structure was beautifully demonstrated in the study of the gradual change of the shape of unit cell of crystalline protein depending upon the water content. This change served as a clue in further analysis of the exceedingly complex diffraction patterns.

The various papers presented at the conference ranged from physics and crystallography to organic chemistry and metallurgy. Among those which may be of particular interest to physicists were the following: J. B. Newkirk, D. L. Martin, and A. H. Geisler described their work on order-disorder phenomena in cobalt-platinum alloys which indicates the presence of a heterogeneous transformation and strong lattice strains. Also in the field of order-disorder belonged the paper by F. W. Lamb and F. Meyer, who found in the $PbCl_2-PbBr_2$ system an order which indicates a preferential substitution of certain particular chlorine atoms by bromine. R. Smoluchowski and R. W. Turner reported an influence of magnetic field on orientation of recrystallized grains in ferromagnetic iron-cobalt alloys, the results pointing to a magnetostrictive effect. The striking optical phenomena related to the domain structure of the ferroelectric $BaTiO_3$ were described and illustrated with a motion picture by N. J. Field, T. E. Catelli, and A. de Bretteville, Jr. R. Pepinsky presented his studies on the application of the Auburn electronic computer to the determination of Fourier transforms which gives, within a fraction of a second, either the scattering function in the reciprocal lattice when the electron density function is known, or vice versa. That precision measurements of lattice constants should be based on precision instruments rather than on various corrective formulas was stressed and the methods necessary to attain precision illustrated by M. E. Straumanis, a well-known authority in that field. An x-ray study of the electron concentration in the covalent bond between the carbon atoms in diamond was described by R. Brill who concluded that in this crystal about $\frac{1}{4}$ of one electron of the outer shell is concentrated in every bond direction. Several papers dealt with the use of Geiger-counter x-ray spectrometers: among them M. Field and M. E. Merchant discussed the use of that instrument for a rapid and accurate study of preferred orientations on surfaces of samples. The important problem of dead time and of its relation to the resulting coincidences and non-linearity of Geiger-counter spectrometers was presented by L. Alexander, E. Kummer, and H. P. Klug, who made a careful comparison of experiment with the theoretical expectations. S. S. Sidhu described the resonance phenomena observed in scattering of x-rays and neutrons by titanium and by titanium carbide and their dependence on the properties of the two radiations. The complicated question of particle size determination from scattering of soft x-rays was discussed by K. L. Yudowitch who drew particular attention to the shape of the particle and the use of long wave-lengths.

At the conference particular attention was attracted by the exhibit of modern x-ray and other scientific equipment by various American and one British firm. The conference, which was made possible through the cooperation of various Pittsburgh educational and industrial organizations, was organized by a committee with Professor Smoluchowski as general chairman of the conference and Professor Klug as program chairman.*

* In addition to the four papers directly following this summary, a Letter to the Editor describing work which was part of this Conference appears on page 816, and the paper by Yudowitch has already appeared in the February issue.

The Precision Determination of Lattice Constants by the Powder and Rotating Crystal Methods and Applications*

M. E. Straumanis
University of Missouri, School of Mines and Metallurgy, Rolla, Missouri
(Received December 9, 1948)

A method that makes possible the attainment of a very high degree of precision in the determination of lattice constants (in some cases more than 1:200,000, even with crystals of lower symmetry) is described. The method and its applications, which has been in steady development since 1934 (in Riga, University of Latvia), tries to eliminate the errors of the Debye-Scherrer-Hull and rotating crystal methods by a careful experimental technique, based on:

(1) Very exactly built cameras of a reasonable diameter (64 mm);

(2) Elimination of all well-known errors of the powder and rotating crystal methods (e.g., elimination of film shrinkage errors by a new method of putting the film into the camera);

(3) Locating the camera in a specially constructed thermostat so that the temperature some hours before and during exposure is kept constant.

The films obtained in this way are very clear, especially in the important back-reflection region, with distinct lines (or spots) that can be measured exactly. As no standard substances are necessary and the absorption corrections are negligible (using very thin samples), the method must be regarded absolute. Of course, the sample (crystal) must be prepared, mounted, and centered under the microscope, a procedure that is not difficult.

Applications: lattice constants and expansion coefficients (even within the range of 10°C); quantitative solubility determinations; appearance of new phases and transition points; the determination of the exact value of Avogadro's number; molecular and atomic weight determinations (if the exact density values are known); statement of imperfections in crystals.

INTRODUCTION

THERE are two ways to attain precision in the determination of lattice constants by the powder and rotating crystal methods, by calculation and by careful experimental technique.

It is possible to eliminate the different errors that are connected with the determination of lattice constants to a certain degree by calculation. Cohen claims that the method described by him is a perfect one, as the absolute values of the constants can be determined by using his equations.[1,2] The precision of this method has already been discussed, in 1936.[3] Some work in this direction was recently done by Weisz, Cochran, and Cole.[4] It must be established for pure physical reasons that measurements of highest precision can be attained only by careful work, eliminating experimental errors as much as possible, and using the calculation methods only when no other solution exists. It is shown in the subsequent sections how these errors can be eliminated.

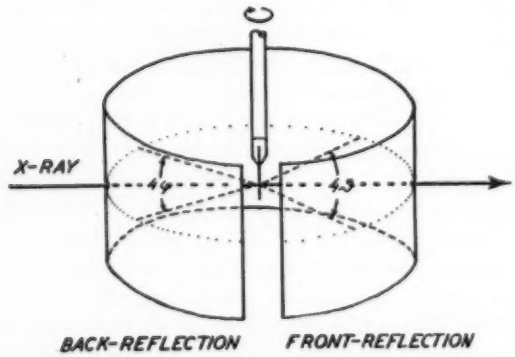


FIG. 1. The new method of loading the camera. ϑ is the Bragg reflection angle, φ the back reflection angle; $\varphi = 90 - \vartheta$ and $\sin\vartheta = \cos\varphi$.

FILM SHRINKAGE, SAMPLE ABSORPTION ERRORS, AND THEIR ELIMINATION. FILM MEASUREMENT

One of the principal errors sourced in lattice constant determination was the unprecise knowledge of the radius of the rolled film as it lies in the form of a cylinder in the cylindrical camera. After it has been developed and dried, the film shrinks gradually with the passing of time, and precise calculation of the reflection angles becomes impossible. Partial elimination of these errors is possible through the use of special cameras and standard substances. These involve new inconveniences.

All these difficulties can now be avoided by the use of a new method of loading the camera. The method was developed in Riga, Latvia, in 1934.[5] The idea is illustrated in Fig. 1.

This scheme is known in this country also as one can find it in M. J. Buerger's book, *X-ray Crystallography*.[6] Figure 2 shows the unrolled film with front- and back-reflection lines on it side by side.

In comparison with former powder diffraction patterns the film has an asymmetric appearance. It is this quality that caused some authors[7] to call this method "asymmetric."

---

* Paper presented to the Sixth Annual Pittsburgh Conference on X-Ray and Electron Diffraction (November 19, 1948).

[1] M. U. Cohen, Rev. Sci. Inst. **6**, 68–74 (1935), Zeits. f. Krist. **94**, 388, 306 (1936).

[2] M. U. Cohen, Rev. Sci. Inst. **7**, 155 (1936).

[3] A. Ievinš and M. Straumanis, Zeits. f. Krist. **94**, 40 (1936); **95**, 451 (1936).

[4] Weisz, Cochran, and Cole, Acta Crysta. **1**, 83 (1948).

[5] M. Straumanis and A. Ievinš, Naturwiss. **23**, 833 (1935); Zeits. f. Physik **98**, 461 (1936). Detailed description of the method: M. Straumanis and A. Ievinš, *Die Präzisionsbestimmung von Gitterkonstanten nach der asymmetrischen Methode*, (Verlag. Julius Springer, Berlin, Germany, 1940), reprinted by Edwards Brothers, Inc., Ann Arbor, Michigan (1948).

[6] M. J. Buerger, *X-ray Crystallography* (John Wiley & Sons, Inc., New York, 1942), p. 395; see also M. J. Buerger, J. App. Phys. **16**, 501, 504, 505 (1945).

[7] F. Halla and H. Mark, *Röntgenographische Untersuchung von Kristallen* (Leipzig, 1937), p. 177.

This method of loading the camera can be used for rotating crystal photographs as well. Figure 3 shows an asymmetric rotating-crystal diagram. From such a diagram the entire space reciprocal lattice can be easily constructed and the lattice type recognized. There is no need for additional photographs with the crystal rotated about other axes, even in some cases for crystals of lower symmetry.[8] For precision determination of lattice constants of a single crystal only the equator (zero layer line) can be used. Photographs can be made, therefore, on narrow film strips (3 cm wide). The advantages of this asymmetric film are:

(1) The knowledge of the camera's radius is unnecessary;

(2) complete independence of film thickness and shrinkage as the precise reflection angles can be calculated from the films themselves;

(3) smaller, normally built cameras can be used.

This is possible because the centers of the front- and back-reflection line systems can be measured on the film with high accuracy at any time. This distance is exactly 180°. Table I shows some reflection angles calculated from a film which was measured 3 days and then 2 weeks after its development.

Accuracy of reflection angle measurements lies between 0.01 and 0.001 degrees. Normal accurately built powder or rotating crystal cameras can be used.

The second source of error in lattice constant determination is the absorption of the sample. Very many formulas and methods have been developed for the exclusion or computation of these errors. There are 3 possibilities of reducing the shift of the lines due to absorption:

(1) using the lines in the back-reflection, sensitive region,

(2) decreasing the diameter of the sample, and

(3) making it more transparent to the x-rays.

The first point is already used by the asymmetric method of loading the camera, which gives the possibility of exact measurement of the lines in that region. The two last points can be achieved by thin samples, not more than 0.2 mm in diameter. The correction, $\Delta\vartheta$, caclulated by Pauli's equation[9] is small in this case,

TABLE I. Reflection angles, ϑ_{mm} measured 3 days and then 2 weeks after film development; Δ_{mm} difference in mm; $\vartheta°$ the angles in degrees; Afga-Laue film; substance TlCl.

Index	Time after development	$2\vartheta_{mm}$	Δ_{mm}	$\vartheta°$
123α	3 days	49.062	−0.097	48.711
123α	2 weeks	48.965		48.721
$224\alpha_1$	3 days	79.752	−0.177	79.182
$224\alpha_1$	2 weeks	79.575		79.185
$224\alpha_2$	3 days	80.572	−0.177	80.000
$224\alpha_2$	2 weeks	80.395		80.001

0.2-mm sample

ϑ	$\Delta\vartheta$
70°	0.006°
80°	0.0015°.

This shift diminishes further, if the samples are more transparent to the x-rays. This increased transparency can be easily obtained by sticking the powder (sieved through a screen of 10,000 meshes/cm²) to the outside of a short piece of practically non-absorbing beryllium-lithium-boron glass hair having a diameter of approximately 0.08 mm. The powder mount thus has a diameter of 0.15 to 0.2 mm. A thin layer of Fisher's *Cello-Seal* can be used for glue (see Fig. 4).

It could be proved that such powder samples behaved quite translucent to x-rays,[10] since the lattice constant of tungsten calculated from the last φ-angles, without using any correction, agree very well with those obtained by Cohen's method and by the method of Jette and Foote: $a=3.1583kX$ (appr. 20°C),[10] 3.14835 (Cohen),[2] and 3.15837 (Jette and Foote, 25°C).[11] Therefore, it is of no use to apply difficult correction methods if the true lattice constants can be obtained directly from the φ-angle measurements. For this purpose all the sharp last lines (in the region of 80° and over) can be used. If there are no lines, the radiation must be changed. However, even with no other lines, the last doublet gives very good results (see Tables IV and V).

Of course, when the samples are thicker, absorption corrections must be introduced or any of the graphical

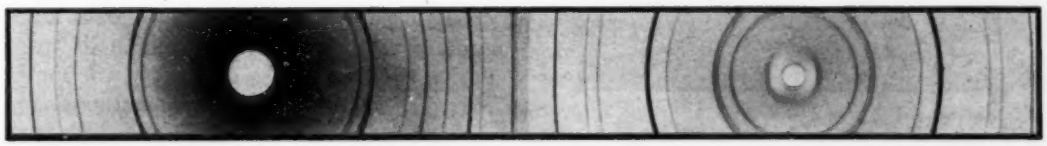


FIG. 2. Asymmetric diffraction pattern film; substance: tungsten; camera diameter: 57.4 mm. The β-lines can also be used for lattice constant measurements.

[8] M. Straumanis, Zeits. f. Krist. **104**, 18 (1942); **104**, 167 (1942). M. Straumanis and E. Ence, Zeits. f. Krist. **104**, 375 (1942)—monoclinic crystals; **105**, 1 (1943)—rhombohedral and triclinic crystals.
[9] O. Pauli, Zeits. f. Krist. **56**, 591 (1921); approximately the same results are given by the Hadding equation.
[10] M. Straumanis and A. Ievins, Zeits. f. Physik. **98**, 461, 470 (1936).
[11] E. R. Jette and R. Foote, J. Chem. Phys. **3**, 605, 615 (1935); Am. Inst. Tech. Pub. No. 670 (1936).

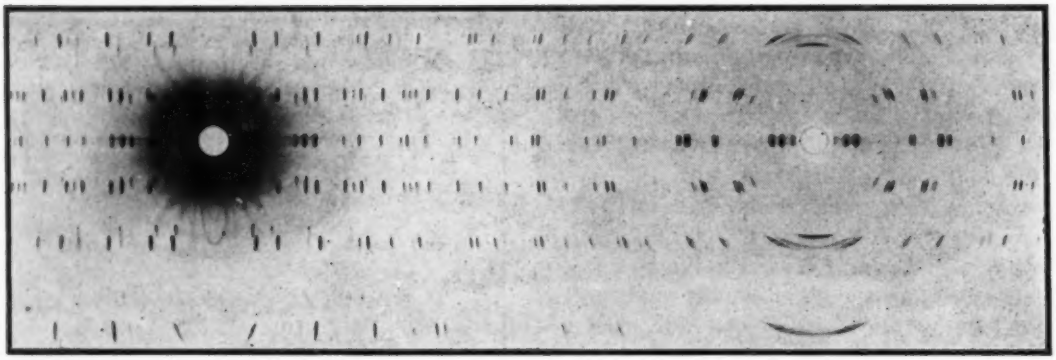


FIG. 3. Asymmetric rotating crystal diagram; substance: single crystal of $NaClO_3$; rotation axis: [001]; camera diameter: 64 mm. In the sensitive back-reflection region 4 doublets can be distinguished (right side). Using a magnifier, 4 more may be seen on the original film.

extrapolation methods can be applied with success.[12] This can be done also if the lines are broad, weak, etc.

Only thin crystals are applicable for rotating crystal photographs used for precision determination of lattice parameters without absorption correction. If they are thicker or the photographs are to be made in other than the needle directions, a certain amount of preparation of crystals is necessary. This can be accomplished by splitting, breaking, sawing, dissolving, and so on. Small special tools are used, mostly under the microscope, and, if necessary, a "micro-manipulator." If the crystal pieces are too tiny to set conveniently by optical methods, some rotating crystal photographs will help overcome the difficulties.[13] In nearly every case sharp lines or spots can be obtained for accurate measurement of reflection angles. However, the rotating crystal method gives slightly lower constants than the power method.[14] Some crystals prepared for exposure are shown in Fig. 5.[15,16]

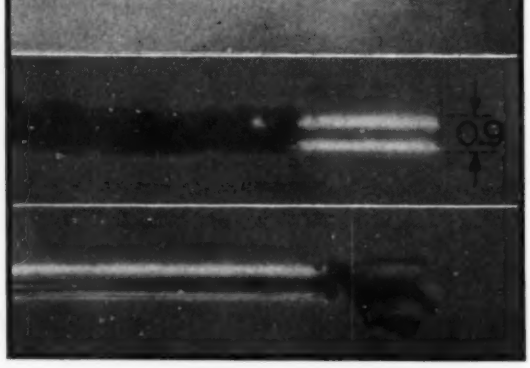


FIG. 4. A 0.12-mm powder mount. (a)—WO_2 on the glass hair (20×); (b)—the same sample (50×), the x-rays can penetrate between and partially through the grains; (c)—a particle of an Ag-Zn alloy on the end of the glass hair (50×).

The exposure time for such tiny crystals or powder samples is longer and increases considerably if the diameter of the sample is less than 0.2 mm. As the grain size of the powder also has some influence on the intensity and smoothness of the lines,[17] sizing through a fine screen (10,000 meshes/cm²) must be recommended.

Mounts of hygroscopic powder (e.g., LiCl, NaBr, etc.)[18] can be obtained by sealing the powder mount into thin-walled glass capillary tubes[19] for obtaining precision powder photographs.

The positions of the last lines must be measured with an accuracy of at least 0.01 mm. With good comparators this is possible. Figure 6 shows a comparator made by David Mann of Lincoln, Massachusetts.

With this instrument one can measure 0.001 mm. Of course, the accuracy is too high for Debye-Scherrer line measurement but it was impossible to find, here in the U. S. A., a convenient instrument of the accuracy needed (0.01 or 0.005).

The possibility of receiving sharp lines in the back-reflection region which are hardly shifted by absorption and the possibility of elimination of shrinkage errors makes the method an absolute one; no standard substances are necessary. The absolute value of the lattice constant can be calculated in *kX* or in angstrom units. And, even if extrapolation methods are applied to derive the true lattice parameter, the method must be regarded as absolute. Other methods, working with calibrating knife edges or with mm scales printed on the

[12] A. J. Bradley and A. H. Jay, Proc. Phys. Soc. London **44**, 563 (1932); J. B. Nelson and D. P. Riley, Proc. Phys. Soc. London **57**, 160 (1945); A. Taylor and H. Sinclair, Proc. Phys. Soc. London **57**, 126 (1945).

[13] O. Weisz and W. F. Cole, J. Sci. Inst. **25**, 213 (1948).

[14] M. Straumanis and A. Ievinš, Zeits. f. Physik **109**, 728 (1938).

[15] M. Straumanis, Zeits. f. Krist. **102**, 432 (1940).

[16] M. Straumanis and J. Sauka, Zeits. f. physik. Chemie, **B53**, 320, 323 (1943).

[17] Alexander, Klug, and Kummer, J. App. Phys. **19**, 742 (1948).

[18] Ievinš, Straumanis, and Karlsons, Zeits. f. physik. Chemie, **B40**, 146 (1938).

[19] Cellulose acetate capillary tubes: K. E. Ben and H. H. Claassen, Rev. Sci. Inst. **19**, 179 (1948).

films, do not possess this property. They need standard substances.

Table II shows how to compute, from a measured film, its present circumference and the angles of reflection.

The figures are the readings of the lines in the comparator. If we add the readings of the corresponding lines around the left hole we obtain a constant number. Any mistake can be detected in this manner. Another constant is obtained from the readings from the lines around the right hole (back-reflection region). These sums agree within some thousandths of a mm. The difference gives at once the present circumference of the film. The difference of the readings of the corresponding lines gives the reflection angles in mm, which can be easily converted into degrees by using the appropriate conversion factor. The smallest φ-angles can be used directly for calculating the lattice constants.[20]

THE THERMOSTAT AND THE PRECISION CAMERA FOR POWDER AND ROTATING CRYSTAL DIAGRAMS. THE SIZE OF THE CAMERA

In order to reach the highest precision a third condition must be fulfilled; it is necessary to work at a constant temperature with the camera in a thermostat.[21] Many authors think it is enough to hold the temperature of the x-ray room constant. Experience shows that this is insufficient. The temperature in the middle of a camera that has been in a thermostat as much as 2 or 3 hours is slightly different from the water temperature of the thermostat. Table III demonstrates how much the lattice constants are affected by temperature changes.

Working at "room temperature," the error due to inaccurate knowledge of temperature lies in the fourth decimal place. Temperature fluctuations of even a few tenths of a degree influence the fifth place. It is clear,

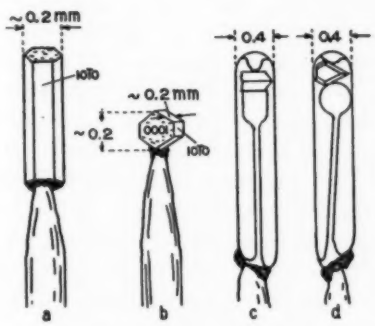


FIG. 5. Tiny crystals mounted on glass hairs for precision determination of lattice parameters by the rotating crystal method. (a) and (b)—tellurium crystals;[15] (c) and (d)—iodine crystals, sealed in thin-walled capillary glass tubes.[16]

[20] Determination of lattice constants of non-cubic substances: see Ievinš, Straumanis, and Karlsons, Zeits. f. physik. Chemie **B40**, 347 (1938); A. Ievinš and M. Straumanis, Zeits. f. Physik **116**, 194 (1940); M. Straumanis and J. Sauka, Zeits. f. physik. Chemie **B51**, 219 (1942).

[21] See also V. Vand, J. App. Phys. **19**, 852 (1948).

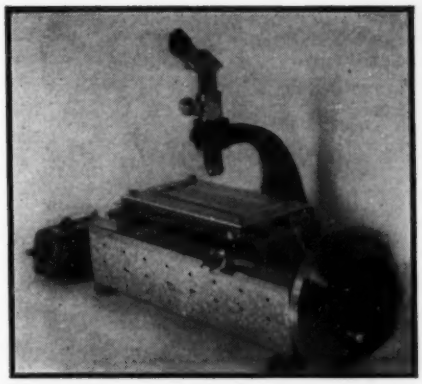


FIG. 6. Comparator built by D. Mann and adjusted for film measurements up to 200 mm length, magnification 3 to 18 ×. The central line of the film can be easily moved into the crosshair of the microscope.

therefore, that the highest precision is only obtained when the loaded camera has been kept at constant temperature for several hours before and during exposure. A section through a suitable thermostat, now being made in Rolla, is shown in Fig. 7.

The camera is placed in the middle of the device. The x-rays enter the camera from the left and leave at the right. The holes around the wall of the camera are water passages through which water at constant temperature flows. The thermostats built in Latvia were fed by a so-called "ultra thermostat" which drove the water through the agency of rotary pumps. The temperature of the flowing water was controlled very exactly (± 0.01°C) by adjustable contact thermometers. Similar thermometers are now available in the U. S. from the Fisher Company.

Of course, cameras for precision measurements must be built very carefully.

(a) The inside of the cylinder of the camera must be exactly round and of the same diameter throughout.

(b) The sample must be in the middle of the camera (maximum deviation ± 0.02 mm).

(c) The sample must be centered and adjusted exactly (maximum deviation ± 0.02 mm).

A camera that corresponds to these requirements and can be used for either powder or rotating crystal photographs is shown in section in Fig. 8.

The rotating shaft *e* can be connected by means of a screw to a small goniometer head (that could not be found in this country) or to a simplified head for the powder sample. Any portion of the sample can be brought into the x-ray by displacing the vertical axis and fastening it with the screw *d*. If there is very little powder available (say 0.1 or 0.01 mg), it can be stuck to the point of a glass needle and moved into the x-ray beam, Fig. 4(c). This may also be done with tiny crystals. The sample can be observed through the collimator with a magnifier. The aperture of the collimator is round (1 mm) as slots give lower results.[5] Figure 9 shows the sample holders for powder photographs.

TABLE II. Calculation of front- and back-reflection angles ϑ and φ from a measured asymmetric film; substance Na_2WO_4; Cu target.

Front-reflection lines	v. w.	v. st.	w.	v. st.	w.	v. st.
mm	64.744	65.641	70.145	71.661	72.595	74.385
	47.914	47.008	42.515	40.988	40.055	38.272
Sum	112.658	0.649	0.660	0.649	0.650	0.657
Diff.	16.830	18.633	27.630	30.773	32.540	$36.113 = 4\vartheta$

Average sum $= 112.654 \pm 0.006$ mm

Back-reflection lines	m α_1	w. α_2	st. α_1	w. α_2	v. st. α_1	m α_2
mm	172.953	172.378	167.682	166.782	162.839	161.129
	139.513	140.084	144.784	145.703	149.638	151.371
Sum	312.466	0.462	0.468	0.485	0.477	(0.500)
Diff.	33.440	32.294	22.898	21.079	13.201	$9.758 = 4\varphi$

Average sum $= 312.472 \pm 0.011$ mm

Circumference $= 312.472 - 112.654 = 199.818$ mm

1 mm on the film $= 360° : 199.818 = 1.80164°$ or $400g : 199.818 = 2.00182g$*

Factor: $1.80164 : 4 = 0.45041$.

* g means "new degrees," $2\pi = 400g$.

The bare film is pressed to the inside of the cylinder with two special rings, one upper and one lower (Fig. 10).

The holes are bored through the film after it is loaded into the camera. This is done from the outside through the holes provided for the beam system.

In order to ascertain whether the axis of rotation of the shaft, *e*, coincides exactly with the axis of the cylinder, *f* (Fig. 8), the arrangement shown in Fig. 11 is used. The holder with a glass hair on its tip is screwed to *e*. The hair is brought into the axis of rotation of *e* as closely as possible and the screw, *d*, is tightened. By turning the cover of the camera with the glass hair thus mounted and by observing the hair through the microscope, one can discover whether or not the hair is exactly in the center of the cylinder. If it does not displace during rotation it is correctly adjusted. Deviations of ± 0.02 mm from the center influence the measurements to a very slight degree. Such deviations, therefore, are permissible. If the sample moves in larger circles, the screws *h* must be loosened and the bearing *k* brought into the right position. Figure 12 shows the camera on an adjustable support, ready for moving into the x-ray beam.

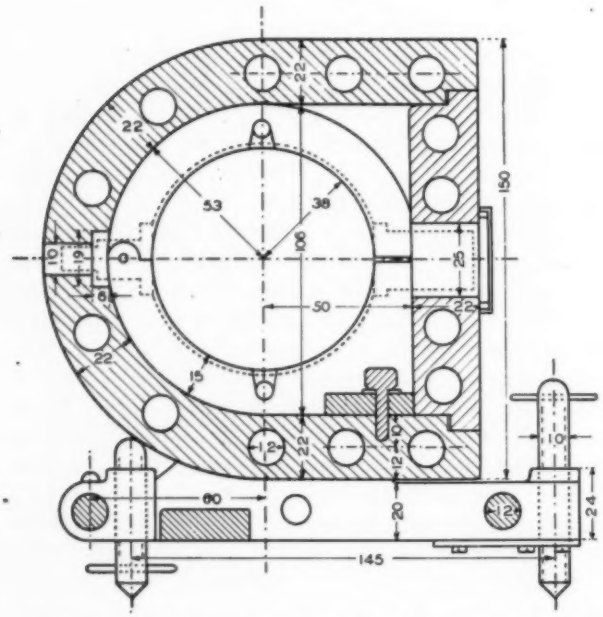


FIG. 7. Section through a thermostat for x-ray cameras. The inside broken line indicates the camera. The door of the thermostat is to the right. The thermostat may be turned around its central axis. It possesses all facilities for easily moving the camera into the x-ray beam. Dimensions are in mm.

This camera was made in Rolla by Professor Kilpatrick of the Department of Mechanical Engineering. It is precision machined throughout.

What size of camera is best for precision measurements? Does a large camera work better than a small one? To decide this important question, seven precision cameras were built in Latvia and the lattice constant of pure aluminum (99.998 percent) was determined in each of them, at constant temperature in thermostats, of course. The results are shown in Table IV.[22]

It follows from this table that too-small cameras (e.g., 28.9 mm) do not work satisfactorily. They are also very inconvenient. Large cameras have no advantage, as the precision of determination is the same as in cameras of middle size. The 57.7-mm cameras are the best in most respects but they are slightly too small if rotating crystal photographs are to be made. However, a small goniometer head fits very well into a 64-mm camera, and this size is to be regarded as being the best. Here 2^g ("new degrees," "Neugrad"$-g$) corresponds to 1 mm (2π equals 400^g).

Why the large cameras do not work better can be explained by examining the films by a recording microphotometer. The lines obtained with a 114.8-mm camera are nearly twice as broad as those obtained with a 57.7-mm camera.[23] The beams reflected by the crystalline substance are not parallal, but divergent. Therefore, there is no advantage in working with large cameras. The resolution is not sharper. Because of the following reasons it is much better to work with precise 64-mm cameras and use the microscope for centering and a comparator for film measurement:

(1) The time of exposure is much shorter in small cameras.

(2) Short films are more uniform and more convenient for measurement than long ones.

(3) A uniform temperature can be maintained better in small cameras than in large ones. Therefore, they are more precise.

(4) A thermostat for a small camera, as well as all additional devices, is small, handier, and cheaper.

[22] A. Ievinš and M. Straumanis, Zeits. f. physik. Chemie **34**, 402 (1936).

[23] A. Ievinš, Zeits. f. Physik **112**, 350 (1939).

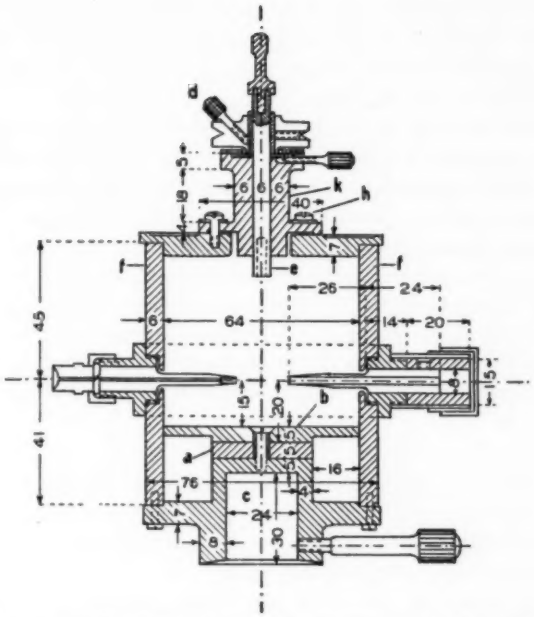


FIG. 8. Section through a precision camera for powder and rotating crystal photographs. As drawn here, films 3 cm in width fit into the camera. Four-cm films will fit if *a* is removed. By removing the bottom, *b*, 8-cm films may be used (see Fig. 3). Originally cameras of this type were designed for a Seemann x-ray apparatus.

The only advantage of large cameras is that the films can be measured with a meter stick and there is no need for expensive comparators.

THE RESULTS OBTAINED WITH THE DESCRIBED METHOD AND CAMERA

From Table IV follows the range of reproducibility of such lattice constant measurements. The highest deviation from the mean value is $\pm 0.00002kX$. The reproducibility is higher than 1:200,000. Usually the diffraction patterns were obtained at different sample temperatures and the calculated constants were then corrected to a standard temperature (18, 20, or 25°C) using the expansion coefficient determined from the just calculated lattice constants. How the constants agree can be seen from the examples of Al and LiF in Table V.

Although there is little agreement in the application of the refraction correction,[24] all the constants below are corrected for this error, using the formula mentioned by Jette and Foote.[11]

In both bases the average deviation from the arithmetic mean is approximately 0.00002, so that the precision 1:200,000 is reached here also. Such a high degree of accuracy could be obtained with other, even noncubic, substances as well. It seems unlikely that the precision of a good spectrograph should be reached with a small 64-mm camera. However, it must not be forgotten that for calculation of lattice constants only

TABLE III. Lattide constants and temperature; α = linear expansion coefficient.

Substance	a_{25} in kX	$\alpha \times 10^6$	Δ_a per 1°C
TlCl	3.83459	54.57	0.00021
Pb	4.94006	29.10	0.00015
Al	4.04145	23.13	0.00008

TABLE IV. Diameter of the camera and the lattice constant of Al.

Camera diameter in mm	Lattice constant in kX corrected to 25°C
114.8	4.04143±0.00002
86	7±0.00001
64	4±0.00002
57.7	6±0.00002
57.7	5±0.00001
57.7	6±0.00002
28.9	18±0.00003
Average except the last constant	4.04145±0.00002

the last lines in the back-reflection region are used. There the resolution is much higher than in the region available to spectrographs (up to 50°). If the last lines can be measured with an accuracy of 0.01 mm, the spectrographic lines must be measured with an accuracy of 0.002 mm. That lies just on the limit of possibility.

In Table VI some of the lattice constants determined by the method being described are compared with the constants of Jette and Foote,[11] Gebert,[25] and Klug.[26]

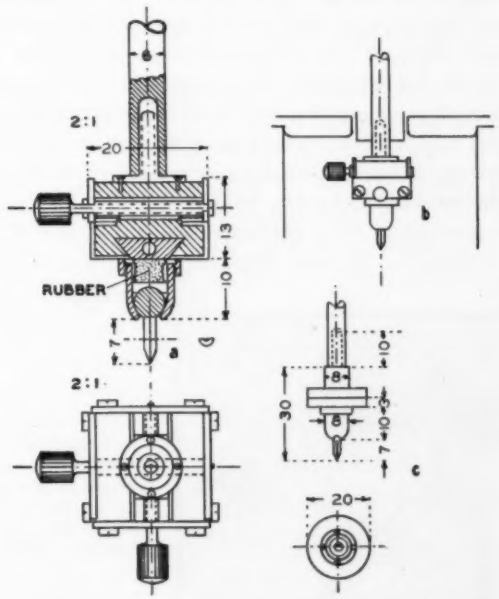


FIG. 9. Adjustable sample holders. The tip, *a*, is provided for the glass-hair-powder mount; *b* shows the holder placed into the camera (Fig. 8); *c* illustrates a magnetic holder, with two gliding planes held together with magnetic forces. These holders allow the whole length of the glass hair to be set into the exact axis of rotation.

[24] E. A. Owen and E. L. Yates, Phil. Mag. VII 15, 472 (1933); A. J. C. Wilson, Proc. Camb. Phil. Soc. 36, 485 (1940).

[25] E. R. Jette and E. B. Gebert, J. Chem. Phys. 1, 753 (1933).
[26] H. P. Klug, J. Am. Chem. Soc. 68, 1493 (1946).

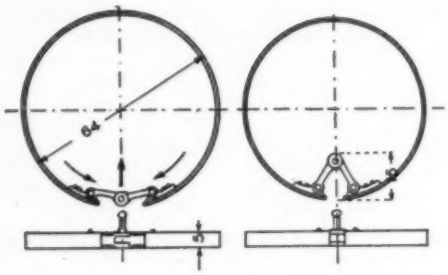


FIG. 10. Rings, closed and opened. They can be covered with rubber. If in the camera the rings can be opened with a small tool.

The lattice constants obtained for aluminum and lead are slightly higher than the values of Jette and Foote; for lead it is lower than the constant of Klug. The agreement is good. Kossel was able to determine, with his spectrographic method, the lattice constants of some crystals with a high degree of accuracy.[27] For the lattice constant of copper he found $3.60751 \pm 0.00004 kX$. The author's earlier results, working with Mellis, were 3.6075.[28] However, there are also some disagreements: It was possible by Kossel's method, requiring large crystals, to reproduce the values of Siegbahn for the lattice constants of rock salt and calcite. Determination of these same constants by the method under discussion gave lower results by 0.006 percent and 0.014 percent.

Although the accuracy of the method itself is high, nothing can be declared about the absolute value of the constants measured. Of course, one can compute the error limit of the absolute determinations using the formula $\Delta d/d = \tan\varphi \Delta\varphi$.[29] but there is another way to test the value of the constants obtained. If the exact density of the substance is known, the x-ray density, or the molecular weight, or the number of molecules in the unit cell, or Avogadro's number can be calculated. Comparing these values with the best ones obtained by other methods, we can judge the values of the constants measured. It is preferable to compute the Avogadro number, as the Siegbahn Avogadro number, $N_s = 6.0594 \times 10^{23}$, must be reproduced from every lattice constant if the calculations were made using Siegbahn's wave-lengths, and if there are no imperfections in the crystals.[30]

The left part of Table VII shows that, on an average, the Siegbahn number is obtained. The lattice constants of these substances, therefore, are the true constants within the error limits. That cannot be said about the lattice parameters of the elements of the right side of the table, where the average does not agree with N_s and the deviations are 10 times greater. It is probable that the densities are not determined accurately enough. Further careful experimental work should lead to the determination of the true values of the lattice constants.

APPLICATION

The usual applications of the powder and the rotating crystal methods are: (a) determination of solubility limits, (b) determination of composition, and (c) determination of changes in structure.

All three applications should be briefly discussed, using sodium-tungsten bronzes for an example. These have metallic properties, such as luster, very good electrical conductivity, and are soluble only in strong oxidizing agents, although they are neither alloys nor intermetallic compounds. The pure bronze has the formula $NaWO_3$ with pentavalent tungsten.[31] $NaWO_3$ absorbs WO_3, forming mixed crystals. $NaWO_3$ is cubic, WO_3, triclinic. In Fig. 13 the solubility of WO_3 in $NaWO_3$ is shown.[32]

It was necessary here to determine the composition of the mixed crystals. Because of the lattice contraction with increasing WO_3 content it was readily possible. The changes in WO_3 content could be determined with an accuracy of 0.1 to 0.05 percent, although the work was not done in a thermostat. How the points experimentally found and calculated agree, can be seen from Table VIII for the first part of the straight line in Fig. 13.

At 28 percent WO_3 there is a transition point in the curve. The crystal absorbs WO_3 without lattice shrinkage. Also at this point the bronzes have the highest

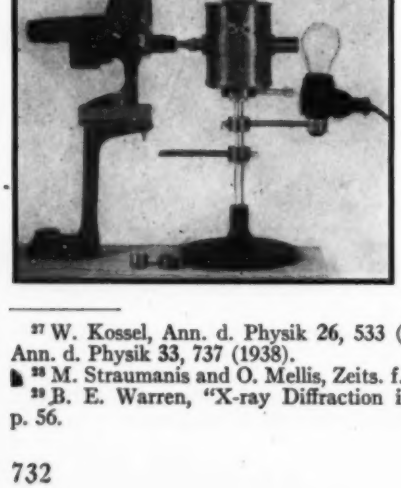


FIG. 11. Arrangement for determining whether or not the sample is in the center of the cylinder. This work must be done only once.

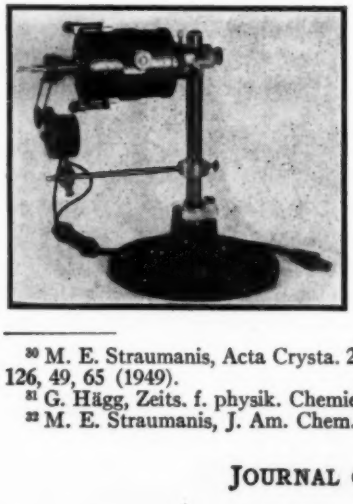


FIG. 12. The camera on the support, which, once adjusted and fastened, makes possible the replacement of the camera in the x-ray beam a matter of a few seconds. The motor is on the left side.

[27] W. Kossel, Ann. d. Physik **26**, 533 (1936); N. van Bergen, Ann. d. Physik **33**, 737 (1938).
[28] M. Straumanis and O. Mellis, Zeits. f. Physik **94**, 184 (1935).
[29] B. E. Warren, "X-ray Diffraction in Crystals," (lectures) p. 56.
[30] M. E. Straumanis, Acta Crysta. **2**, 82 (1949); Zeits. f. Physik **126**, 49, 65 (1949).
[31] G. Hägg, Zeits. f. physik. Chemie **B29**, 201 (1935).
[32] M. E. Straumanis, J. Am. Chem. Soc. **71**, 679 (1949).

TABLE V. Lattice constants of pure Al (99.998%) and LiF.

Film No.	t°C	Lattice constant in kX Measured	Corrected to 25°C
596	25.60	4.04149	4.04144
597	25.60	153	8
599	25.85	154	9
599*	25.85	151	7
598	44.30	327	6
		Average = 4.04147 kX ± 0.00003	
659	26.38	4.01819	4.01800
664	26.34	1829	9
668	46.22	2097	5
673	59.44	2276	2
675	14.57	1662	5
800	26.48	1828	8
		Average = 4.01805 kX ± 0.00005	
		Refraction 0.00003	
		4.01808	

* This measurement of the film was repeated after 3 weeks.

TABLE VI. Lattice constants in kX of some cubic (a) and non-cubic substances (a and c) at 25°C. in comparison with the constants of Jette, Foote, Gebert, and Klug (Δ).

	a	Δ	c	Δ
Al	4.04145	+0.00006	—	—
Pb	4.94006	+0.00146*	—	—
Pb	4.94006	−0.00074**	—	—
Sn	5.81970	+0.00020	3.17488	−0.00012
Bi	4.53674	−0.00046	11.83834	+0.00024
Mg	3.20280	−0.00020	5.19983	−0.00038

* Jette and Gebert. ** Klug.

electrical conductivity, comparable with that of graphite.[33] In measurements of minor accuracy this point could not be detected. The observer would probably draw a curve through the mean of all the observed points.

A further application is the determination of expansion coefficients, especially of substances of lower symmetry. This can hardly be accomplished by other methods.[34] Selenium has two linear expansion coefficients, one in the "a" direction (α_1) the other in the direction of the "c" axis (α_2).[35]

The table shows that in the "a" direction the expansion coefficient is positive, while in the "c" direction it is negative, that is, the "c" dimension of the lattice contracts with increasing temperature.

Then, as already mentioned, the Avogadro number can be determined. However, we cannot expect any other number than Siegbahn's ($N_s = 6.0594 \times 10^{23}$) on which the wave-length measurements are based. Nevertheless, using the correction factor $k = 1.0002$ (due to the raising of the molecular weight of calcite

[33] M. E. Straumanis and A. Dravnieks, J. Am. Chem. Soc. **71**, 683 (1949).
[34] M. Straumanis and A. Ievins, Zeits. f. anorg. Chemie **238**, 175 (1938).
[35] M. Straumanis, Zeits. f. Krist. **102**, 432 (1940).

TABLE VII. The Siegbahn Avogadro number, N_s, from lattice constants and densities.

Substance	$N_s \times 10^{-23}$	Substance	$N_s \times 10^{-23}$
Ag	6.0603	Al	6.055
Cu	593	Au	84
Fe	586	Bi	74
Pb	605	I	65
Te	591	Mg	70
LiF	596	Se	65
$Pb(NO_3)_2$	584	Sn	63
Average	6.0594 ± 0.0011		6.066 ± 0.015

TABLE VIII. The deviation of the measured points from linearity; calculated lattice constants for the first 6 points: $a_{calc} = 3.8519 - 0.000829 \times$ percent WO_3.

Wt. WO_3 %	a meas. in kX	a calc in kX	Difference	Difference %
12.08	3.8422	3.8419	+0.0003	+0.0078
13.96	3.8403	3.8403	±0.0000	±0.000
16.88	3.8380	3.8379	+0.0001	+0.0026
21.88	3.8335	3.8338	−0.0003	−0.0078
24.5	3.8312	3.8314	−0.0002	−0.0026
26.48	3.8308	3.8300	+0.0003	+0.0078

from 100.075 to 100.095, the most recent figure, including mixed crystal formation)[30] and using the wave-length conversion factor 1.00203 or 1.00202, we can calculate the absolute Abogadro number N_0.

$$N_0 = \frac{1.0002 \times 6.0594 \times 10^{23}}{(1.00203)^3}$$

$$= 6.02385 \times 10^{23} \quad \text{or} \quad 6.02403 \times 10^{23}.$$

Of positive value is the determination of atomic and molecular weights of high accuracy, first carried out by Johnston, Hutchison, and Hutchison.[36] By a different

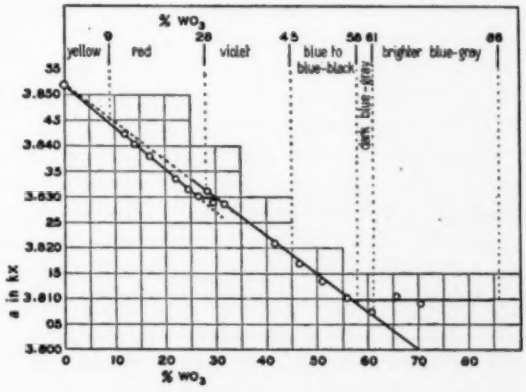


FIG. 13. The system $NaWO_3$–WO_3. The solubility limit of WO_3 in cubic $NaWO_3$ is between 56 and 57 percent WO_3 by weight. The color changes gradually with the WO_3 content. Transition zone between 27 and 30 percent WO_3. The content of WO_3 can be calculated from the lattice contraction (see Table VIII).

[36] C. A. Hutchison and H. L. Johnston, J. Am. Chem. Soc. **62**, 3165 (1940); **63**, 1580 (1941); H. L. Johnston and D. A. Hutchison, Phys. Rev. **62**, 32 (1942); C. A. Hutchison, J. Chem. Phys. **10**, 489 (1942); D. A. Hutchison, J. Chem. Phys. **13**, 383 (1945).

TABLE IX. The linear expansion coefficients α_1 and α_2 of Se (trigonal).

t °C	a_t kX	$\alpha_1 \times 10^4$	t °C	c_t kX	$\alpha_2 \times 10^4$
40.00	4.35985	73.30	35.00	4.94805	−17.89
20.10	350		15.00	982	
60.10	4.36641	74.86	55.00	4.94628	−17.89
40.00	5985		35.00	805	
60.10	4.36641	74.10	55.00	4.94628	−17.89
20.10	5350		15.00	982	
	Average	74.09			−17.89

TABLE X. Atomic and molecular weights of some elements and compounds determined by means of x-rays.

Substance	Atomic weight chem.	x-rays	Δ	Δ %
Ag	107.88	107.86	−0.02	−0.019
Cu	63.54	63.540	±0.000	±0.000
Fe	55.85	55.858	+0.008	+0.014
Pb	207.21	207.17	−0.04	−0.019
Te	127.61	127.62	+0.01	+0.0078
F*	19.00	18.999	−0.001	−0.0052
LiF	25.940	25.939	−0.001	−0.0039
$Pb(NO_3)_2$	331.226	331.282	+0.056	+0.017
Calcite	100.095	100.065	−0.03	−0.03

* $M_{LiF} - M_{Li} = M_F$.

method (using the Avogadro number), a series of atomic and molecular weights, using the densities from the literature, and the lattice constants determined by the method described, were calculated.[30] The best determinations are summarized in Table X.

The difference between the chemical weight (M) and the weight by means of x-rays (M_x) is very small. There is a hope that this method will make it possible to determine more accurate atomic weights for some elements, especially those whose weights it is difficult to find by chemical methods (Mg, Al, Sn, Bi, Au, and others). Even molecular weights of complicated organic substances can be determined in this way.[37]

Finally, it may happen that the M_x, although the greatest care was taken for their precise determination, do not agree with the corresponding chemical constants within error limits. In this case, and only in this case, have we the right to speak about imperfections in crystals.

If $M_x - M = -D$, there are vacant sites; if $M_x - M = D$, there are interstitial atoms and if $M_x - M = 0$, the crystal is sound.

The accuracy of determination at present is 3 imperfections in 10,000 sites. Only in the event that the number of imperfect sites is determined as higher than 3 may we speak about imperfections in crystals.[30]

And so with the precision determination of lattice constants and densities of the substances examined, the circle is closed: the atomic and molecular weights found by chemical methods, the weights calculated from x-ray data and experimental densities, the lattice constants, the x-ray densities, and the Avogadro's number can be mutually related very exactly. From the possible differences in these relations (e.g., $M_x \neq M$) we can judge the perfection of the crystals.

ACKNOWLEDGMENT

I sincerely thank Professor A. V. Kilpatrick of the Mechanical Engineering Department of Missouri School of Mines and Metallurgy for the careful work he did in making the camera and other equipment shown in the figures of this article.

[37] W. C. Bunn, *Chemical Crystallography*, Oxford, Clarendon Press (1945), p. 185; see also Palmer, Ballantyne, and Galvin, J. Am. Chem. Soc. **70**, 906 (1948).

Dead Time and Non-Linearity Characteristics of the Geiger-Counter X-Ray Spectrometer*

LEROY ALEXANDER, ELIZABETH KUMMER, AND HAROLD P. KLUG
Department of Research in Chemical Physics, Mellon Institute, Pittsburgh, Pennsylvania
(Received February 7, 1949)

The satisfactory use of Geiger-counter methods for the measurement of diffracted x-ray intensities demands that the observed counting rates be corrected for the non-linear response of the counter. Both the multiple foil method of calibrating the non-linearity of response and the two-source method of Beers for measuring counter dead time have been found to be unsatisfactory approaches to this problem. The experience of this laboratory indicates that electronically controlled oscillographic techniques for the direct measurement of dead time are much more satisfactory. With a knowledge of the dead time, τ, the true counting rate, N, can be calculated from the observed counting rate, N_o, by employing the formula $N=N_o/(1-N_o\tau)$.

For a pulsating x-ray source the effective counting rate is faster than the observed by a factor which is characteristic of the type of rectification employed, the target material, and the voltage applied to the x-ray tube. A sufficiently accurate value of this factor can be arrived at by oscillographic observation of the pulse pattern and in some cases by mathematical analysis.

I. INTRODUCTION

THERE is a general awareness among users of the Geiger-counter spectrometer that accurate intensity measurements can be made only if the observed counting rates are corrected for the non-linear response of the counter. However, there is general disagreement as to the magnitude and importance of the corrections and as to the best method for effecting them. Thus some laboratories report apparent linearity of response for the Norelco spectrometer up to nearly 1000 counts per second while other users of the same instrument find the need for correction of counting rates as low as 100 to 150. It seems doubtful that spectrometers of very similar and often identical design can manifest actual response differences as great as these.

During the past two years the authors have pursued a study of quantitative analytical procedures using the Norelco x-ray spectrometer, and occasion has been found for testing and comparing several of the present-day methods for non-linear response corrections. The present paper is principally a summary of this experience, and it is presented in the hope that it will assist in dispelling some of the uncertainty that has shrouded this subject.

II. PRELIMINARY DISCUSSION

Non-linearity of response of a Geiger counter means that the counting rate is not proportional to the intensity of the incident radiation. Two ways in which this deficiency of response may be detected are as follows: (1) If the incident x-radiation of wave-length λ is reduced in intensity by passage through a series of absorbing foils of uniform thickness x, the removal of one foil increases the intensity of the emergent radiation by the constant factor $I_{n-1}/I_n=e^{\mu x}$, μ being the linear absorption coefficient of the foils for the monochromatic x-rays. However, the ratio of the counting rates, N_{n-1}/N_n, does not remain constant, but shows a falling off as n diminishes and the intensity increases. (2) If two radioactive sources are counted separately and then together, the counting rate for the combined sources is observed to be less than the sum of the counting rates of the separate sources.

In the usual self-quenching Geiger-counter circuits the non-linearity of response originates largely in the finite resolving time of the Geiger tube. In the present paper it is assumed that coincidence errors due to the time constants of the amplifier circuit may be neglected. In addition the finite resolving time of the mechanical register does not cause appreciable errors where a sufficiently large scaling ratio is used. As shown by Alaoglu and Smith,[1] losses due to the mechanical register vanish when $n\geq\sigma/\tau$, where n is the scaling ratio, and σ and τ are respectively the resolving times of the mechanical register and Geiger counter. As this condition was either maintained or closely approached at all times during the present study, the register was disregarded as a source of counting losses.

Coincidence losses are actually related to the resolving time, rather than the dead time, of the Geiger tube. The resolving time is defined as the time which must elapse following a pulse before a succeeding pulse will have sufficient amplitude to trip the first stage of the scaling circuit. Evidently, then, the resolving time varies somewhat with the degree of amplification preceding the first scaling stage. Because of this indeterminancy, and because the resolving time is only slightly larger than the dead time in usual practice, the two terms will be used synonymously in the present discussion.

If τ is the dead time and N_o is the observed counting rate, the true counting rate is given by

$$N=N_o/(1-N_o\tau). \quad (1)$$

This formula is quite accurate for corrections up to

* Based on material presented before the Sixth Annual Pittsburgh Conference on X-Ray and Electron Diffraction at Pittsburgh, Pennsylvania (November 20, 1948).

[1] L. Alaoglu and N. M. Smith, Phys. Rev. 53, 832 (1938).

6 percent and probably accurate enough for the present investigation when the correction is of the order of 10 percent.

III. MULTIPLE FOIL METHOD FOR THE DETERMINATION OF NON-LINEAR RESPONSE CORRECTIONS

This method has been used by Lonsdale[2] in measuring spectrometrically the intensities of reflections from single crystals. The theoretical basis for this procedure is as follows. Monochromatic x-rays of wave-length λ are reduced in intensity by a series of absorbing foils of uniform thickness x and linear absorption coefficient μ before being received by the Geiger tube. If the response were completely linear, the relative counting rates for n and 0 foils would be equal to the relative intensities, or

$$N_n/N_o=\epsilon^{-n\mu x} \tag{2}$$

and

$$\log_{10}N_n=\log_{10}N_o-(\mu x/2.303)n. \tag{3}$$

This equation shows that in the absence of coincidence errors a plot of $\log_{10}N_n$ against n, the number of foils, should give a straight line with a slope of $-\mu x/2.303$.

Figure 1 shows how coincidence corrections and the dead time can in theory be determined by this method. The solid line is a smooth curve drawn through the experimentally observed counting rates for various numbers of foils ranging from 0 up to a number which is sufficient to reduce the counting rate to a very low level, for example, 5 counts per second. The beam intensity must be great enough to give an intensity with 0 foils which is at least as great as the largest which is anticipated in actual work. Following the plotting of the experimental points, the theoretical curve is precisely located by drawing a straight line with slope $-\mu x/2.303$ which passes through the experimental points at very low counting rates where the response correction is negligibly small.

These curves can now be used to correct succeeding experimental measurements for response deficiency in

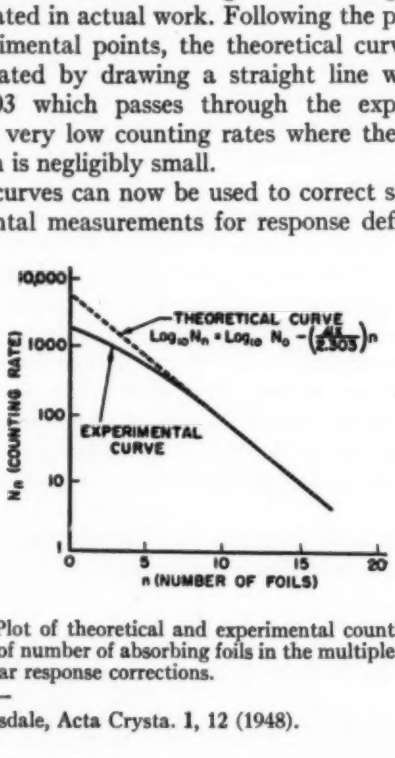


FIG. 1. Plot of theoretical and experimental counting rates as a function of number of absorbing foils in the multiple foil method for nonlinear response corrections.

either of two ways: (a) By direct reference to the graph one can read the theoretical counting rate, N_n, for any observed counting rate, $(N_n)_o$. This method is applicable to the highest counting rates which have previously been calibrated. (b) One can calculate the dead time by using Eq. (1) in the form

$$\tau=\frac{N_n-(N_n)_o}{N_n(N_n)_o}, \tag{4}$$

after which Eq. (1) is used to correct observed counting rates. This method is less satisfactory than (a) because it is not applicable when the corrections are much greater than 6 percent. However, it has the special advantage of supplying a value of the dead time, which may be compared with values arrived at by other methods.

Although the multiple foil method of calibration is sound in principle, it has not proven satisfactory in practice for a number of reasons. First and most important, the foil thickness is seldom accurately known. This makes the slope of the ideal line uncertain, and one is forced to resort to the device of extrapolating from experimental points at low counting rates with no means of allowing for the small but appreciable response corrections at these points. This results in a significant error in the extrapolated values of N_n at higher counting rates. Second, there is generally a rather large variation in thickness from foil to foil, which causes large uncertainty limits in the extrapolated values of N_n. Third, statistical counting errors always must be reckoned with, particularly at low intensities where very long counting periods are required. Fourth, even though a crystal-diffracted ray is utilized, the radiation is not truly monochromatic owing to the presence in the general radiation of second and third order harmonics of the characteristic wave-length. Lonsdale[2] has treated this harmonic problem in some detail and shows that for CuKα characteristic radiation it is negligible for nickel but relatively serious for aluminum foil. Nickel is particularly suitable because of its relatively high absorption coefficients for the general radiation harmonics, which reduces their contribution to the total diffracted energy to a very low percentage even at counting rates as low as one per second.

Figure 2 illustrates the uncertainty in the extrapolation method. By taking particular pains, the probable errors in counting rate due to foil thickness variations and to counting statistics can be reduced to ± 1 percent each or to a combined effect of $\pm(1^2+1^2)^{\frac{1}{2}}=\pm 1.4$ percent. It is seen that an extrapolation between the points $N_1=30$ and $N_2=150$ to the point $N_3=750$ sec.$^{-1}$ leads to an uncertainty in N_3 which is three times as great as that in either N_1 or N_2, or ± 4.2 percent. Besides this degree of uncertainty in the extrapolated point N_3, it is also subject to a systematic negative error which arises from the appreciably greater coincidence loss

[2] K. Lonsdale, Acta Crysta. 1, 12 (1948).

at the point N_2 than at N_1. Thus, for the typical case of $\tau=150$ microseconds, the observed values of N_1 and N_2 will be less than the theoretical values (30 and 150 sec.$^{-1}$) by about 0.5 and 2.3 percent, respectively, as calculated with Eq. (1). Hence, N_2 will be approximately 2 percent low in relation to N_1, and this will extrapolate as a -4 percent error in N_3.

These sources of error have a telling effect upon the accuracy of response corrections and dead time determinations performed by the graphical method. Suppose, for example, that the dead time of the Geiger tube is 150 microseconds and that an extrapolation is to be carried out from the points N_1 and N_2 to the point N_3 (Fig. 2). Disregarding for the moment the uncertainty due to foil thickness variations and counting statistics, the effect of the systematic error discussed above will be to impose a -4 percent error on N_3, giving an *apparent* theoretical counting rate of 720 as compared with the true value of 750 sec.$^{-1}$. In addition, the effect of foil thickness variations and counting statistics imposes an uncertainty of ± 4.2 percent or ± 30 counts per second. This leads to limiting values of the *apparent* theoretical counting rate of 720 ± 30, or 690 and 750 sec.$^{-1}$. But since we know that the true theoretical counting rate is 750, we may use Eq. (1) to calculate the average counting rate which should be observed, which is found to be $(N_3)_o=674$ sec.$^{-1}$. By means of Eq. (4) we then find for the upper and lower limiting values of the dead time as determined by the graphical method:

$$\tau_{min}=(690-674)/(690\times 674)=34\ \mu\text{sec.}$$
$$\tau_{max}=(750-674)/(750\times 674)=150\ \mu\text{sec.}$$

These results are seen to differ by nearly fivefold.

The foregoing analysis of errors is only approximate in nature, but it is based on typical experimental data and, therefore, should predict correctly the general magnitude and nature of the errors to be expected. Evidently the dead time determined by the multiple foil method may be expected to vary all the way from approximately the true value down to very small values which are greatly in error.

These expectations have been confirmed by multiple-foil measurements of the dead times of two Norelco Geiger-counter tubes, Types 62002 (Lindeman glass window) and 62003 (mica window). In order to reduce systematic and random errors to a practicable minimum, the following features of procedure were observed: (1) The Norelco x-ray unit and counting circuits were allowed to warm up for two hours before measurements were made. (2) The power input to the equipment was regulated to ± 0.2 volt by means of a Sorensen Model 1000 electronic regulator and additionally by manual adjustment if necessary. (3) A diffracted beam of ample energy was obtained by diffracting CuKα radiation from the $(10\bar{1}1)$ planes of a quartz plate. (4) In order to preclude the possibility of a drift from the position of maximum intensity of the reflection, several counts

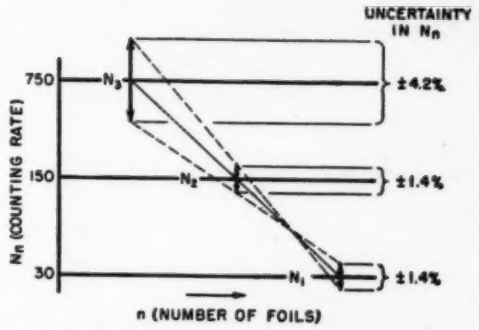


FIG. 2. Uncertainty in extrapolating the theoretical curve in the multiple foil method for nonlinear response corrections.

were always taken across the $(10\bar{1}1)$ maximum at 0.025° intervals, thus delineating the crest of the contour. (5) In order to compensate for any possible variations in the electrical characteristics of the circuits, a measurement of N_o, the counting rate with 0 absorbing foils, was made at frequent intervals and a correction factor applied to the measurements to put them on a constant basis. Although the data were taken over a period of several days, this factor never fell outside the limits 0.97–1.03, and it usually lay between 0.98 and 1.02. (6) Electrolytic nickel foil of 0.00035-in. mean thickness was used, the variation in thickness being within ± 10 percent according to the manufacturer. The maximum number of foils used was 15 and the minimum number 0 (except for the standard nickel filter supplied with the instrument, which was never removed). In order to reduce the effect of thickness variations, measurements of at least 13 randomly chosen foils were averaged in arriving at a best value of N_n for each value of n. Thus, for various representative values of n the following numbers of randomly chosen foils were used: for $n=15$, one set of 15; for $n=13$, one set of 13; for $n=12$, two sets of 12; for $n=8$, two sets of 8; for $n=5$, three sets of 5; and so on. A simple analysis shows that this reduces the probable error limits in N_n due to foil thickness variations to about ± 1 percent. With 0 and 15 foils, high and low apparent counting rates of respectively about 1000 and 10 per second were obtained. (7) Statistical counting errors were kept below a probable error of ± 1 percent by recording at least 5000 counts for each measurement.

The observed counting rates were plotted logarithmically against the number of foils in the manner shown in Fig. 1, and the best straight line was drawn through the 5 or 6 points in the experimental curve which fell below an apparent counting rate of 75. The dead time was then calculated from several pairs of N_n and $(N_n)_o$ values for which the correction did not greatly exceed 6 percent. It should also be mentioned that the observed counting rates were all multiplied by a factor of three before the plotting was performed. The reason for this correction will be discussed in detail in Part V, and it suffices here to say that it is necessitated by the pulsating character of the x-ray source.

TABLE I. Dead time measurements on Norelco Geiger-counter tubes. All values in microseconds; Geiger tube voltage 1400.

	Oscillographic method I	II	Average	Multiple foil method
Type 62002 (Lindemann glass window)	136	155	146±10	96
Type 62003 (mica window)	168	167	168±10	149

Table I gives the results obtained for the two Norelco tubes by the multiple foil method and compares them with the more dependable values subsequently obtained by the oscillographic method of Stever.[3] It is seen that the results by the foil method bear no consistent relationship to the more accurate values except that both are smaller as predicted by the analysis of errors above.

IV. TWO-SOURCE STATISTICAL METHOD OF BEERS[4,5] FOR THE DETERMINATION OF DEAD TIME

In this method two radioactive sources are counted separately and then combined, giving counting rates N_1, N_2, and N_{12}. The dead time is then calculated from the formula

$$\tau=(N_1+N_2-N_{12})/2N_1N_2. \quad (5)$$

Beers found that consistent results were obtained by this method only if pure β-emitters were used. γ-emitters and mixed emitters tended to give non-reproducible results. Ghelardi[6] at Tracerlab reports similar difficulties with non-β-emitters.

This method is subject to other difficulties. In dry atmospheres spurious counts may be obtained due to the accumulation of charges on the Geiger tube window, a phenomenon which would be expected to become more serious with long counting periods. Also long counting periods are required since about 200,000 and 400,000 counts must be totalized for the separate and combined sources respectively in order to reduce the probable error in the dead time due to counting statistics to ±10 microseconds.

Table II shows the variability of results obtained by the authors in 10 determinations of the dead time of a Norelco Geiger tube Type 62002. In each test counting rates of about 100 and 200 sec.$^{-1}$ were used respectively for the separate and combined sources. Measurements 1 and 2 were made using one radium and one x-ray source, an incorrect procedure in view of the fact that a radium source is a constant statistical emitter whereas the x-ray source emits photons in 60-cycle bursts.

It is probably significant that when shorter counting periods were tried, measurements 9 and 10, the results were smaller and approached more nearly the more reliable oscillographic result of 146 microseconds. However, statistical counting errors are much higher in these cases so that it is mostly coincidental that the two results agree exactly. These results bear out the findings of Beers and Ghelardi that γ-emitters lead to inconsistent results. Evidently the method is unreliable for the determination of the dead time of a Geiger tube designed for the detection of x-ray quanta.

V. OSCILLOGRAPHIC METHODS FOR THE DIRECT MEASUREMENT OF DEAD TIME

Of the various rccognized methods for response corrections, electronically controlled oscillographic techniques for the direct measurement of dead time appear to be the most satisfactory. This laboratory has successfully applied the method of Stever[3] using the one-shot sweep circuit suggested by Ghelardi.[6] Pulses from the preamplifier of the Geiger tube are used to trigger a one-shot sweep circuit and in addition are fed to the vertical plates of an oscilloscope. The output of the sweep circuit is supplied to the horizontal plates. Since each horizontal sweep is synchronized with the application of a pulse to the vertical plates, a strong persistent pulse pattern appears on the screen. If the sweep time is somewhat greater than the dead time, and if the counting rate is moderately high, a statistical array of fainter pulse contours is observed on the screen near the end of the sweep. The distance on the screen from the beginning of the initial pulse contour to the position at which the faint pulses begin to appear is a measure of the dead time. By calibrating the sweep time with an audio oscillator of known output frequency, the measured dead time distance can be converted to time in appropriate units.

Ghelardi suggests that an oscilloscope of good deflection sensitivity be used, such as the Du Mont Model 274, and that the pulses from the Geiger tube and the output of the one-shot sweep circuit be fed directly to the oscilloscope plates. However, the authors have found it advantageous to employ a Du Mont Type 208-B oscilloscope, which reproduces the incoming signal with such negligible distortion that the horizontal and vertical amplifiers and gain controls can be used in the normal way. This greatly facilitates carrying out the calibration and measurements.

Table I gives dead times for two Norelco Geiger tubes, Types 62002 and 62003, as determined by the method of Stever. Measurements I and II were in each case performed by different persons and it is seen that they are in reasonably good agreement. It is believed that the probable error of the average result is not greater than 10 microseconds.

Pepinsky and co-workers[7] have devised an oscillographic method which appears to constitute a refinement of Stever's method and should be capable of somewhat higher accuracy. In this procedure the x-ray

[3] H. G. Stever, Phys. Rev. **61**, 38 (1942).
[4] Y. Beers, Rev. Sci. Inst. **13**, 72 (1942).
[5] Tracerlog No. 3, p. 4 (April, 1947).
[6] R. P. Ghelardi, Tracerlog No. 6, p. 5 (October, 1947).
[7] Pepinsky, Jarmotz, Long, and Sayre, Rev. Sci. Inst. **19**, 51 (1948).

tube is pulsed into conduction by two sets of ½-microsecond, 1000-cycle triggers which are continuously phasable with respect to each other. The two sets of x-ray pulses produce two Geiger pulse contours on the oscilloscope screen, and by reducing the time separation (phase difference) until the second pulse contour disappears the dead time can be directly determined. By this method Pepinsky found for a Norelco Type 62002 Geiger tube dead times of 100 and 140 microseconds respectively at applied voltages of 1500 and 1250. This is in approximate but not close agreement with the value of 146 microseconds for a tube of the same type measured in the present study by Stever's method.

VI. COINCIDENCE LOSSES IN COUNTING A PULSATING X-RAY SOURCE

In Part III it was pointed out in connection with the evaluation of spectrometer measurements that the observed counting rates should be multiplied by a factor of three to allow for the pulsating character of the x-ray source. This correction factor is only approximately right and was arrived at in the following manner. Figure 3A shows the momentary appearance of the oscillographic pattern when Geiger pulses produced by the radiation from a radium source are applied to the vertical plates and the horizontal sweep is set at 60 cycles per second. Figure 3B shows the pattern obtained when 36-kv diffracted x-rays from the Norelco spectrometer source produced the Geiger pulses, which again were synchronized with an x-axis sweep frequency of 60 cycles. This time the pulses appear in clusters with a 60-cycle frequency. The diagram shows how the pulses occur only during the phase of each cycle when the K x-ray spectrum is being generated (only during the negative portion of the cycle and when the negative potential exceeds the critical excitation potential for the K spectrum). Measurements of this pattern of pulse bursts showed that pulses were being produced during only about one third of the total elapsed time.

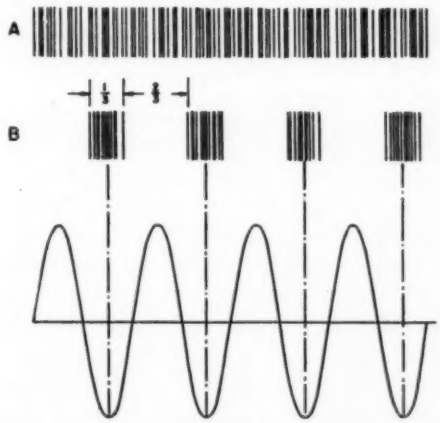


FIG. 3. Oscillographic patterns of Geiger pulses from A, a radioactive source, and B, a self-rectifying x-ray source showing the relationship between the pulse bursts and the phase of the potential applied to the x-ray tube.

TABLE II. Variability of dead time as determined by the statistical method of Beers. Norelco Geiger-counter tube Type 62002.

Measurement number	Source 1	Source 2	Minutes counted	Calculated dead time (microseconds)
1	Ra (solid)	x-ray	40	579
2	Ra (solid)	x-ray	40	462
3	Ra (solid)	Ra (solution)	40	514
4	Ra (solid)	Ra (solution)	40	306
5	Ra (solid)	Ra (solution)	40	365
6	Ra (solid)	Ra (solution)	40	negative
7	Ra (solid)	Ra (solution)	40	576
8	Ra (solid)	Ra (solid)	40	900
9	Ra (solid)	Ra (solid)	10	228
10	Ra (solid)	Ra (solid)	5	228

Hence the true counting rate is approximately three times the apparent one. Since the distribution of pulses within each cluster is not statistical, this factor of three is not really correct, but it nevertheless is sufficiently accurate to serve for the correction of coincidence errors in any experimental work except the most exacting intensity measurements. The size of the correction factor will be expected to vary with the type of rectification employed, the target material, and the voltage applied to the x-ray tube. Hence, it is desirable that the factor be independently determined by each investigator for his own particular x-ray generator and specific experimental conditions.

If the x-ray intensity is a known periodic function of time, an expression for the coincidence loss can be derived as follows. Let the theoretical and observed counting rates at time t be N_t and N_{ob} respectively, where $N_t=f(t)$, $f(t)$ being a periodic function of time with a period that is a submultiple of unity. The theoretical and observed numbers of counts received in time dt at the time t are then, respectively, $f(t)dt$ and $N_{ob}dt$. But from Eq. (1) we know that for corrections of not much over 6 percent we can write with considerable accuracy:

$$N_{ob}=N_t/(1+N_t\tau)=f(t)/(1+f(t)\tau). \tag{6}$$

The net theoretical and observed counting rates (counts theoretically and actually obtained in unit time) will, therefore, be respectively

$$N=\int_0^1 f(t)dt \tag{7}$$

and

$$N_o=\int_0^1 \frac{f(t)dt}{1+f(t)\tau}, \tag{8}$$

and the coincidence loss is found to be

$$N-N_o=\int_0^1 \frac{f^2(t)\tau dt}{1+f(t)\tau}. \tag{9}$$

We now wish to determine for what theoretical and observed counting rates, N' and N_o' respectively, the

same coincidence loss would be suffered if the x-ray source were non-pulsating (completely random emission of photons). In general N' and N_o' will be larger than N and N_o by some factor k. That is to say, $N'=kN$ and $N_o'=kN_o$. But from Eq. (1) we know that $N_o'=N'/(1+N'\tau)$ and hence we can write

$$N_o=N'/[1+N(k\tau)]. \tag{10}$$

Solution of Eq. (10) for k gives

$$k=(N-N_o)/NN_o\tau. \tag{11}$$

By evaluating Eqs. (7) and (8) for the appropriate form of the function $f(t)$ and knowing the dead time τ, one can use Eq. (11) to calculate the correct value of the factor k by which observed counting rates should be multiplied before employing Eq. (1) to determine the theoretical counting rates. Equation (10) is subject to the special interpretation that $k\tau$ is an effective dead time for theoretical and observed net counting rates of N and N_o from a pulsating source. In other words, counting rates from pulsating sources can be corrected for coincidence losses with the aid of Eq. (1) in either of two ways: (a) The observed counting rates may be first multiplied by the proper factor k and the true dead time τ used, or (b) the observed rates may be substituted directly in the equation and the effective dead time $k\tau$ used.

By way of illustration we shall evaluate k for the case of a self-rectifying x-ray tube such as is incorporated in the Norelco spectrometer. We shall assume that the applied potential varies sinusoidally with a 60-cycle frequency, that, as a first approximation, the x-ray intensity varies as the square of the voltage in excess of the critical excitation voltage, $V-V_o$,[8] and that x-rays are produced only during the negative half of each cycle. Then the intensity of the characteristic x-radiation can be expressed as

$$I=K'(V_m \sin 2\pi\nu t-V_o)^2, \tag{12}$$

if it is remembered that $I=0$ when $(V_m \sin 2\pi\nu t-V_o)$ is positive. For CuKα radiation $V_o\cong 9000$ volts, and if we employ a peak value, V_m, of the applied potential of 36,000 volts, we obtain from Eq. (12)

$$I=K'V_o^2(4\sin^2\pi\nu t-1)^2.$$

The theoretical counting rate at time t is proportional to I so that this equation can be written

$$N_t=K(4\sin^2\pi\nu t-1)^2=f(t). \tag{13}$$

Substitution of this value of $f(t)$ in Eqs. (7) and (8) gives for the theoretical and observed net counting rates:

$$N=K\int_0^1(4\sin^2\pi\nu t-1)^2dt \tag{14}$$

and

$$N_o=K\int_0^1\frac{(4\sin^2\pi\nu t-1)^2dt}{1+\tau K(4\sin^2\pi\nu t-1)^2}, \tag{15}$$

the integrands being zero when $(4\sin^2\pi\nu t)<1$.

When Eqs. (14) and (15) are evaluated analytically for representative values of τ and K and a frequency of 60 cycles, the value of the factor k from Eq. (11) is found to be approximately 3.5. For very small values of the product τK (small dead time and/or low counting rates) the value of k converges toward a limiting value of about 3.51, while for the largest values of τK likely to be met in practice k never assumes values smaller than about 3.3. These analytical results are in satisfactory agreement with the oscillographically estimated value of 3.

VII. GENERAL CONCLUSIONS

(1) The multiple foil method for the determination of non-linear response corrections is theoretically capable of high accuracy, but practical difficulties rather severely limit the usefulness of this procedure. An accurate knowledge of the effective foil thickness is needed to make the method satisfactory.

(2) The two-source statistical method of Beers gives inconsistent results when γ- or mixed β- and γ-emitters are used in the calibration. It does not appear to be suitable for calibrating a Geiger-counter x-ray spectrometer.

(3) Direct oscillographic methods of observing the dead time appear to be satisfactory. The one-shot sweep circuit method of Stever has been found to give good results. With a knowledge of the dead time τ and the observed counting rate N_o, the true counting rate is then calculated using Eq. (1), which is sufficiently accurate for corrections at least as high as 6 percent.

(4) In order to allow for possible variations in different samples of the same model of Geiger tube, it is recommended that the dead time of each tube be individually measured.

(5) With anything but a continuous d.c. high voltage x-ray source, the emission of x-rays is discontinuous and consequently the effective counting rate for diffracted x-rays is higher than the apparent one. When correcting observed counting rates for coincidence losses, allowance must be made for this situation by multiplying the observed counting rate by the proper factor. For any given x-ray source this factor can be determined with sufficient accuracy by observing the Geiger pulses with an oscilloscope synchronized with the frequency of the x-ray generator. If reasonable assumptions can be made about the high potential wave form, the factor can be confirmed mathematically.

[8] Charles S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), p. 52.

Reflection Method of Determining Preferred Orientation on the Geiger-Counter Spectrometer

MICHAEL FIELD* AND M. EUGENE MERCHANT
Research Department, The Cincinnati Milling Machine Company, Cincinnati, Ohio
(Received February 28, 1949)

A rapid and accurate method of determining preferred orientation in the surface of thick samples was devised using a reflection technique on the Geiger-counter spectrometer. By presenting the specimen face to all possible angles with respect to the incident x-ray beam, through use of a special fixture, and recording the intensity of diffraction on a potentiometer recorder, the relative number of planes of a given set at each orientation can be determined. The change of intensity due to the positioning of the specimen face away from the Bragg angle is taken care of by a correction formula which is experimentally verified. The method is especially valuable for dealing with thin layers on thick specimens.

THE usual method of evaluating preferred orientation in a metal specimen has been to take a series of transmission photographs of the specimen at different angles in order to obtain sufficient information to construct a pole figure for the low order planes.[1–3] Such a procedure is long and tedious and is subject to various errors.[4]

Decker, Asp, and Harker[5] recently devised a method for obtaining pole figures on transmission specimens using the Geiger-counter x-ray diffraction spectrometer. The method gives intensity readings which can be plotted directly on a polar net to give a pole figure. It is not, however, directly applicable to determining preferred orientation in surface layers of relatively massive specimens (too thick for x-ray transmission measurements). Yet specimens of this type constitute one of the most important classes in practice. These include machined surfaces, surfaces shotpeened or otherwise cold worked for purposes of improving fatigue resistance, and surfaces of other types of massive specimens which have been subjected to special surface treatments. The present paper describes a rapid and accurate method devised for determining pole figures and preferred orientation for such specimens using a reflection arrangement on the Geiger-counter spectrometer.

DETERMINATION OF PREFERRED ORIENTATION ON X-RAY SPECTROMETER

The specimen is positioned in the special fixture shown on the Geiger-counter spectrometer in Fig. 1. The fixture alone, with specimen in place, is shown in Fig. 2. The Z-shaped gauge in the foreground is used to locate the plane of the specimen face exactly on the vertical axis of the fixture. The specimen, Fig. 3a, is mounted so that its face is parallel to the face of the bushing, Fig. 3b. The specimen is held in the bushing by pouring in a molten material of low melting point, such as sulfur or a low melting alloy, and allowing this to solidify. The plate, Fig. 3c, is used to position the specimen in the bushing during the mounting process, to insure that it protrudes by a suitable amount. The fixture itself is so designed that the specimen can be rotated about an axis normal to the specimen face as indicated in the insert of Fig. 1. The axis of rotation is at the same height as the axis of the incident x-ray beam. The fixture can also be swiveled about a vertical axis, i.e., the face of the specimen can be positioned so as to depart from angle θ. An index plate is provided so that α, the angle of departure from θ, can be determined.

In making a preferred orientation study the fixture is first positioned so that the specimen face is at an angle θ with respect to the incident x-ray beam. At this time the Geiger-Mueller tube is at 2θ. The θ, of course, corresponds to a definite (hkl) plane whose preferential orientation is being studied. The specimen is then rotated by means of the synchronous motor and gear reduction of Fig. 2 so that the specimen makes one revolution in $7\frac{1}{2}$ minutes. Meanwhile the intensity

FIG. 1. Geiger counter spectrometer with special fixture for preferred orientation reflection studies.

* Present address: Metcut Research Associates, Cincinnati, Ohio.

[1] F. Wever, Trans. A.I.M.E. **93**, 51–77 (1931).
[2] C. S. Barrett, Trans. A.I.M.E. **124**, 29–58 (1937).
[3] M. Gensemer and R. F. Mehl, Trans. A.I.M.E. **120**, 277–292 (1936).
[4] B. F. Decker, Proc. A.S.T.M. **43**, 785–802 (1943).
[5] Decker, Asp, and Harker, J. App. Phys. **19**, 388–392 (1948).

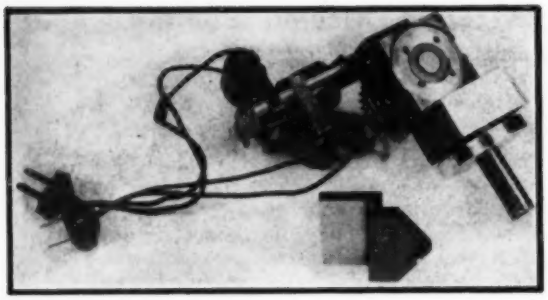


FIG. 2. Special fixture for mounting specimen for preferred orientation studies.

output of the Geiger tube is recorded on the potentiometer recorder normally used in conjunction with this type of x-ray spectrometer. If the grain size is small (and it will be for a worked surface) a uniform intensity is recorded. The fixture is then swiveled away from θ by angle α, but the Geiger tube is held at angle 2θ. The specimen is again rotated one revolution about its own axis. Any preferential orientation of the (hkl) planes at this angle α will result in a variable intensity recorded by the potentiometer recorder. The intensity is, in fact, a measure of the number of (hkl) planes in position to diffract. Since the travel of the recorder chart may be synchronized with the angle of rotation of the specimen, it is possible to correlate the intensity record directly with the angle of rotation of the specimen, known as β. The specimen is then swiveled to a new angle α and the procedure repeated.

Now when passing from one angle of swivel, α, to another, there is a change of intensity even if there is no preferred orientation. This change is caused by several conditions:

(1) The focusing effect of the spectrometer is altered;
(2) the area of the specimen irradiated by the incident x-ray beam is changed;
(3) the relative distances traveled by the incident and the diffracted rays through the specimen vary.

It is therefore necessary to correct for the change in intensity caused by swiveling the sample to the various α's, and the manner in which this is done is described in the succeeding section of this paper. After this correction is made, the intensity recorded for any position of the specimen face can be related to the relative number of (hkl) planes at that same position, and the results plotted directly on a polar net.

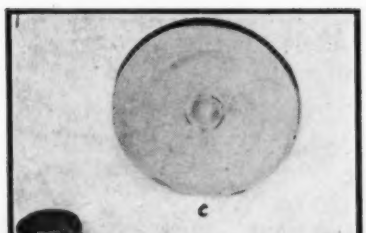


FIG. 3. Specimen mounting bushing, specimen itself, and mounting plate.

CORRECTION FOR CHANGE IN INTENSITY WITH α

The correction for change in intensity of the diffracted beam with α (the departure of the specimen face from θ) is determined as follows: Consider the geometry pictured in Fig. 4.

Let

A = incident intensity,
I_0 = total diffracted intensity when $\alpha=0$,
dI = diffracted intensity from an element of thickness dx in the general case,
D = fraction of incident intensity diffracted at any point in the sample,
x = path length within sample before diffraction occurs,
v = path length within sample after diffraction occurs,
μ = coefficient of absorption of material,
θ = Bragg angle,
α = angle of swivel of specimen surface away from θ. (Positive values of α correspond to counter-clockwise rotation of the fixture as viewed from above, and vice versa.)

The incident intensity after penetration into the specimen a distance x, is $Ae^{-\mu x}$. Then for a diffracting element dx, the diffracting intensity transmitted is

$$dI = Ae^{-\mu x}Ddxe^{-\mu v}. \tag{1}$$

Now, from the geometry of Fig. 4,

$$v = x\sin(\theta+\alpha)/\sin(\theta-\alpha)$$

or

$$v = Mx, \tag{2}$$

where

$$M = \sin(\theta+\alpha)/\sin(\theta-\alpha). \tag{3}$$

Then, from Eqs. (1) and (2)

$$dI = Ae^{-\mu x}Ddxe^{-\mu Mx}$$

or

$$dI = ADe^{-(1+M)\mu x}. \tag{4}$$

To get the total intensity of the diffracted beam, Eq. (4) must be integrated from $x=0$ to $x=\infty$ (specimen considered to be of infinite thickness as regards penetration by the x-ray beam).

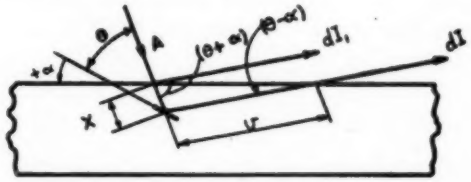


FIG. 4. Geometry of the diffraction process in the reflection specimen.

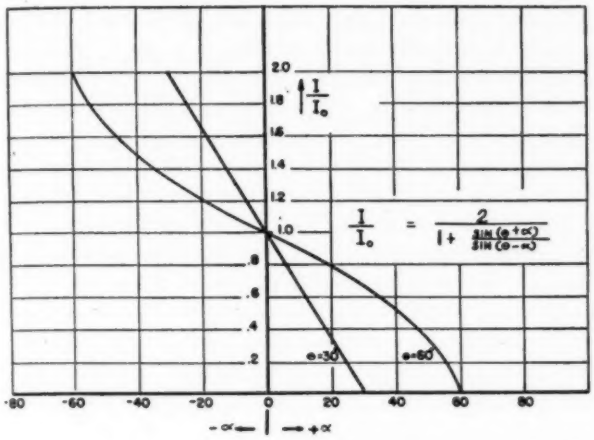


FIG. 5. Plot of Eq. (8) for two different values of Bragg angle.

Then

$$I = AD\int_0^\infty e^{-(1+M)\mu x}dx. \tag{5}$$

Integrating (5) and substituting the limits, gives

$$I = \frac{AD}{(1+M)\mu}. \tag{6}$$

Now, when $\alpha=0$, $M=1$ and therefore

$$I_0 = \frac{AD}{2\mu} \tag{7}$$

and so

$$\frac{I}{I_0} = \frac{2}{1+M}$$

or, from Eq. (3)

$$\frac{I}{I_0} = \frac{2}{1+\dfrac{\sin(\theta+\alpha)}{\sin(\theta-\alpha)}}, \tag{8}$$

which is the equation required for correcting for the change in diffracted intensity with changing α.

Equation (8) is plotted in Fig. 5 for $\theta=30°$ and $\theta=60°$. It is interesting to note that for negative values of α, $I/I_0>1$ or the intensity I increases with respect to I_0. The maximum possible value of I/I_0 is seen to be 2 when the magnitude of $-\alpha$ equals θ. Other facts relating to Eq. (8) should be noted: (1) the ratio of intensities is independent of the mass absorption coefficient—this can be interpreted to mean that the diffracting volume is independent of α; (2) when $\theta=90°$ then $I/I_0=1$ for *all* values of α including $\alpha=\pm 90°$; (3) the magnitude of the swivel angle α can never be greater than θ.

Physically, item 3 means that when the specimen is swivelled to an angle $+\alpha=\theta$, the specimen face cuts off the reflected beam I_0. When the specimen is swivelled to an angle $[-\alpha]=\theta$, then the specimen face cuts off the incident x-ray beam. Theoretically, in order to make a complete preferred orientation study, the entire range of angles from $\alpha=0°$ to $\alpha=90°$ should be explored. Thus, for (*hkl*) planes having θ values considerably less than 90°, the pole figures obtained by the present method are incomplete, lacking some of the outer latitude circles. While this is a limitation of the present method, it is not particularly serious, since

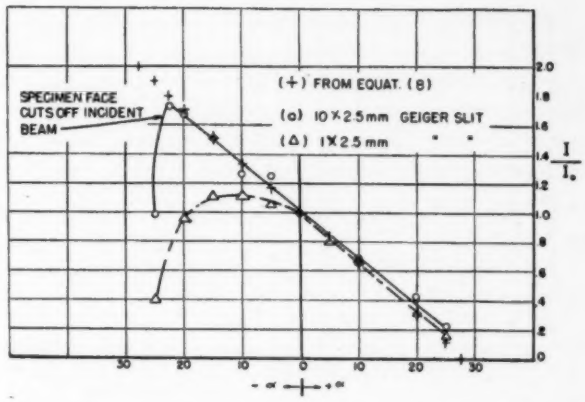


FIG. 6. Experimental verification of validity of Eq. (8) using annealed sample of AISI 302 stainless steel.

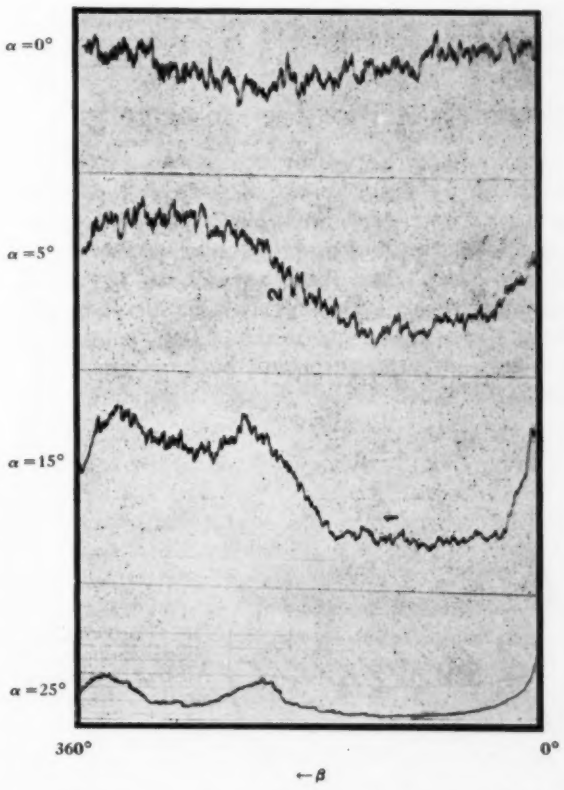


FIG. 7. Intensity records for (111) plane at representative values of swivel angle α, for face milled specimen of AISI 303 stainless steel.

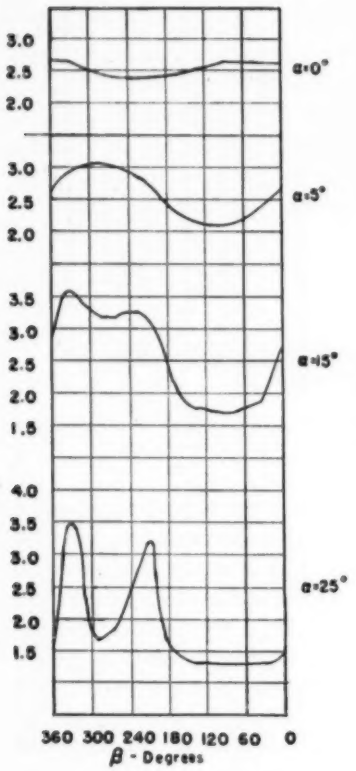


FIG. 8. Plot of intensity values of Fig. 7 after correction by application of Eq. (8). (Relative values.)

the orientation is ordinarily quite well defined by the portion of the pole figure which can be obtained, as is illustrated by the examples given in the last section of this paper.

EXPERIMENTAL VERIFICATION OF EQ. (8)

In order to verify Eq. (8) by experiment, a specimen of A.I.S.I. 302 stainless steel was mounted in the fixture of Fig. 2 with its annealed and electropolished surface exposed to the x-ray beam. $Fe-K_\alpha$ radiation was employed and the Geiger tube was set at $2\theta=55.25°$ corresponding to the (111) plane. The specimen was rotated at 40 revolutions per minute about its own axis so as to minimize the effects of the large grain size. The slit in front of the x-ray tube was adjusted to $\frac{3}{4}$ mm width and 2.5-mm height. The slit height in front of the Geiger tube was maintained at a height of 2.5 mm but the width was varied from a minimum of 1 mm to the maximum possible width of 10 mm. The results are shown in Fig. 6 where I/I_0 is plotted against positive and negative values of α. In the region of positive α, a close correspondence was found between the intensity ratio based on Eq. (8) and the experimental results for the 1-mm wide Geiger slit. The correspondence was poorer in the $+\alpha$ region while using the wide Geiger slit, probably because of the background effects. In the region of negative α there was a poor correspondence between Eq. (8) and the experimental results while employing the 1-mm Geiger slit, because, as the specimen is swivelled in the direction of $-\alpha$, the diffracted beam widens and only a small portion of the beam is received through the narrow slit at the Geiger tube. However, if the 10-mm slit at the Geiger tube is employed, then a close correspondence is obtained in the region of $-\alpha$ until this angle becomes so great that the specimen face cuts off the incident x-ray beam.

The most accurate combination was thus found to be the $\frac{3}{4}$-mm wide x-ray slit and the 1-mm wide Geiger slit and positive values of α; this combination was used in the preferred orientation studies reported here. It was not necessary to use negative values of α, for this region could be explored by setting the sample at $+\alpha$ and rotating to $(\beta+180)$.

RESULTS OF TYPICAL STUDIES

Examples of typical pole figures made by this method are shown in Figs. 9 and 10. These were obtained from preferred orientation studies made on a large number of annealed specimens of AISI 303 stainless steel. The specimens had been subjected to face milling operations using sintered carbide cutters and several different tool angles and cutting speeds. While the results shown in

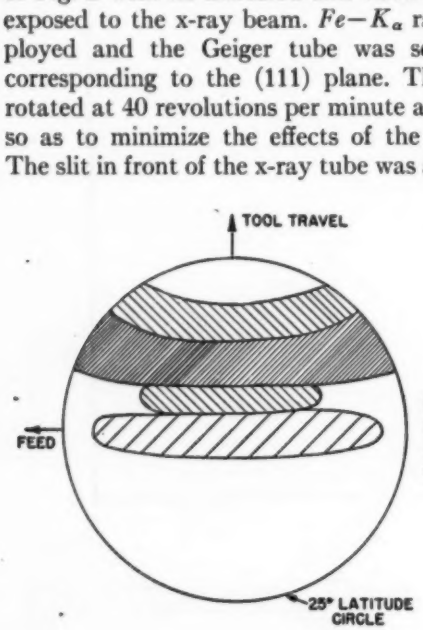


FIG. 9. Pole figure for (111) plane, face milled specimen of AISI 303 stainless steel.

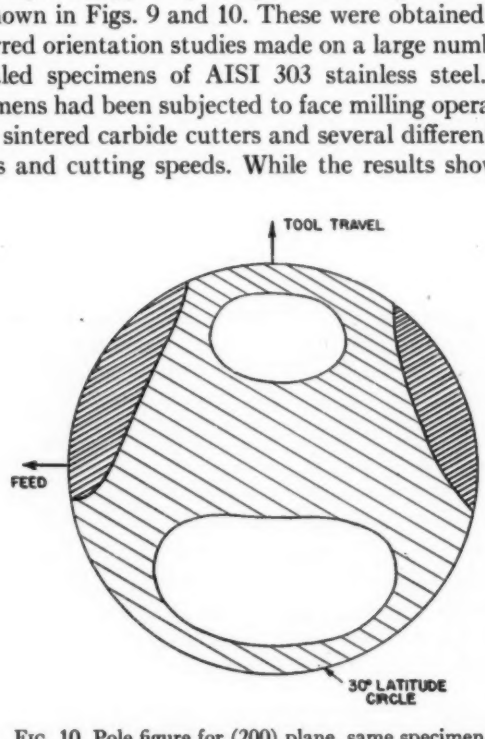


FIG. 10. Pole figure for (200) plane, same specimen as Fig. 9.

Figs. 9 and 10 represent the preferred orientation found in the surface of a particular specimen, the pole figures obtained from the other specimens machined under different conditions were similar.

Fe$-K_\alpha$ radiation was used. Two planes were investigated: the (111) plane for which $\theta=26.6°$ and the (200) plane for which $\theta=32.4°$. The orientation was determined from $\alpha=0°$ to $\alpha=25°$ for the (111) planes and from $\alpha=0°$ to $\alpha=30°$ for the (200) planes. The uncorrected intensity records obtained for the (111) plane at several representative values of α are shown in Fig. 7, and the corrected values of intensity, obtained by application of Eq. (8), are plotted in Fig. 8. The resulting pole figure is that of Fig. 9.

The (111) planes, Fig. 9, were found to be inclined at an angle of 15° to 20° to the milled surface in the direction of the tool travel. Their orientation was symmetrical with the direction of tool travel. The (200) plane orientation, Fig. 10, is in accord with the restrictions set up by the (111) orientation of Fig. 9.

ACKNOWLEDGMENTS

The authors wish to acknowledge, with thanks, the guidance and assistance offered by Professor Harold J. Kersten of the Physics Department, University of Cincinnati and Mr. Hans Ernst and Mr. Norman Zlatin of the Research Department, Cincinnati Milling Machine Company.

Influence of Magnetic Field on Recrystallization

R. Smoluchowski and R. W. Turner
Carnegie Institute of Technology, Pittsburgh, Pennsylvania

(Received March 1, 1949)

In order to gain additional information about the origin of preferred orientation of recrystallized grains, the influence of a magnetic field on recrystallization of iron-cobalt alloys is being studied. Preliminary results indicate a change of the texture which can be qualitatively accounted for by considering the magnetostrictive properties of the alloy and their influence on the stability of the recrystallization centers.

A DEFORMED metal, upon heating, can return to an approximately strain-free condition in two ways: a gradual elimination of strains in such a way that each deformed grain retains its general shape and orientation, or a gradual growth of nearly strain-free grains which usually have an orientation and ultimate shape different from the original grain. In this admittedly oversimplified picture the first process is called recovery, the second recrystallization, and usually both are active at the same time.

Neither recovery nor recrystallization is well understood although both have received much attention recently. This is particularly true of the origin and mechanism of growth of the recrystallization centers.[1] It is doubtful whether the fact that recrystallization at later stages can be described in terms of nucleation and growth[2] justifies the conclusion that the early embryos, or nuclei, of strain-free metal originate and attain a stable condition in a manner similar to usual heterogeneous reactions. The similarity appears to be that of the formal description rather than of the basic mechanism.[3]

In order to provide additional insight into the various factors controlling growth and orientation of recrystallized metals, a study has been undertaken of the influence of a magnetic field on the recrystallization texture. The texture, i.e., preferred orientation of grains, is due to the fact that a deformed metal usually has a texture obtained during deformation and so recrystallization proceeds in an anisotropic medium. It was hoped that a magnetic field might influence the stability of variously oriented recrystallization centers in a ferromagnetic material and thus provide an easily controlled factor in the process of formation of a texture. These expectations proved to be correct and this paper presents a preliminary report of our results.[4]

EXPERIMENTAL PROCEDURE

An obvious choice for our experiments was an iron-cobalt alloy because it combines a high Curie temperature with a relatively low temperature above which, in a deformed sample, recrystallization rather than recovery is the predominant phenomenon. The alloy containing 35 percent cobalt was rolled from about 0.75 mm thickness down to 0.025 mm, at the Westinghouse Research Laboratories, and then annealed in hydrogen at around 700°C in the magnetic field of the large electromagnet at the same laboratories. Several samples were annealed at the same time and completely recrystallized, each having a different orientation of the

[1] M. Cook and T. L. Richards, J. Inst. Metals **73**, 1 (1946); D. Turnbull, "Metals Technology," Tech. Pap. 2365 (June 1948).
[2] W. A. Anderson and R. F. Mehl, Trans. A.I.M.E. **161**, 140 (1945).
[3] R. Smoluchowski, "Theory of Nucleation," National Research Council Conference on Phase Transformations in Solids held at Cornell University (August 1948).
[4] The results were reported at the International Conference on Physics of Metals in Amsterdam (July 1948), and at the Sixth Annual Pittsburgh Conference on X-Ray and Electron Diffraction (November 1948).

direction of rolling, and therefore of the "deformation texture," with respect to the direction of the uniform field. Runs without magnetic field were made in exactly similar conditions except that the electromagnet was

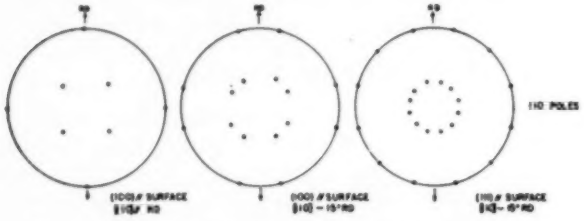


FIG. 1. Ideal (110) pole figures for the three orientations found in recrystallized Fe-Co alloy.

turned off. The preferred orientation present in the samples was then studied by means of an integrating x-ray camera at the Carnegie Institute of Technology. By integrating over an area of about 7 sq. cm, good, continuous diffraction rings were obtained, permitting a proper evaluation of relative intensity. The measurements were made first visually and then checked by means of a photometer.

RESULTS

First the "rolling texture", i.e., the preferred orientation of grain in the rolled alloy, was obtained. It appears to be predominantly the [110] (100) texture in which the [110] direction of the various grains tends to be in the rolling direction and the (100) plane approaches the position parallel to the surface of the sample.

Then the recrystallized texture without magnetic field was studied. It is a rather complicated texture but it can be satisfactorily described as a sum of three ideal simple textures: [110] (100), [110] 15° (100) and [110] 15° (111). Figure 1 shows the 110-pole figures corresponding to these three ideal orientations. The first one is identical with the rolling texture, the other two are characterized by a 15° angle between the [110] direction and the direction of rolling. Similar textures have been reported for other metals.[5]

The influence of the magnetic field appears to be a strong enhancement of the [110] (100) recrystallized texture at the expense of the other two orientations. A semi-quantitative illustration is provided in Fig. 2 in which the shift of the concentration of the [110] poles from the position 15° away from the rolling direction to the position in the rolling direction under the influence of the field can be seen. The great increase of intensity, but not of the spread, of the area in the rolling direction parallel to the field appears clearly in the photometric intensity measurements. A comparison was made of the influence of the magnetic field in the direction of rolling and at right angles to the direction of rolling, but still in the plane of the sample.

[5] See C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943).

The results indicate that the [110] direction in the direction of the field tends to be held parallel to the field, while those at right angles to the field are permitted to rotate about the direction of the field, Fig. 2.

DISCUSSION

The results summarized above show that a magnetic field can influence appreciably the recrystallization

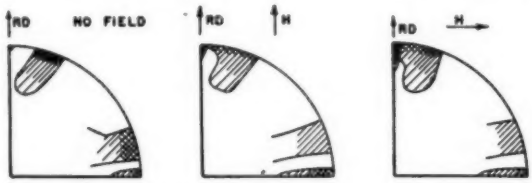


FIG. 2. Comparison of the (110) pole figures obtained without magnetic field, with field parallel to and perpendicular to the rolling direction.

texture. Since these studies were conducted for only one temperature and one rate of initial heating, one may expect to find other conditions which will show a still stronger influence. It appears that the rate of initial heating has an influence upon the intensity of the effect, and this indicates an influence of the magnetic field on the early stages of the recrystallization process.

The results agree in a qualitative manner with the following explanation: Iron-cobalt alloys in the range of composition used here have a high positive magnetostriction in the [100] direction, which is the direction of easy magnetization, and rather low magnetostriction in the [110] direction. Let us compare a small volume of a strain-free metal which has the [100] direction in the direction of the field with another similar volume which has the [110] direction in the direction of the field, and let us assume that in the absence of the field both volumes would have equal chance to grow. If the magnetic field is present, then the volume which has the [100] direction in the direction of the field will try to expand in that direction but, being constrained by the surrounding material, it will become strained. The other volume will have only a small tendency to undergo a magnetostrictive change of shape and thus it will not become strained. As a result of this difference in strain energy the second volume will have a lower free energy, and, hence, it will be more likely to grow than the first one. This agrees with the observations.

There is a host of various questions which arise in connection with these results, and a series of experiments with various alloys, various heat treatments, and various applications of the magnetic field is under way.

The authors wish to express their appreciation of the cooperation of Drs. J. E. Goldman and T. D. Yensen of the Westinghouse Research Laboratories, in making available the alloy used and the facilities for heat-treating in a strong magnetic field. Thanks are also due to Mr. A. Kimmel, of the Carnegie Institute of Technology, for his careful work in plotting the pole figures.

The Morphology of Carbon Black Particles in Shadow-Cast Specimens

JOHN H. L. WATSON
Edsel B. Ford Institute for Medical Research, Detroit, Michigan
(Received November 9, 1948)

Observations are given which favor a reticulate chain structure for carbon black. The discovery of "hollow" particles in carbon black dispersions of P-33 and Thermax is reported along with possible explanations for the observed image intensities which give rise to the interpretation of hollowness. Striated surfaces strongly suggestive of a laminar structure and serrated edges which conform in orientation with these plateaus have been detected. Rough areas arranged both in lines and at random are observed over the surfaces of some particles. Shape variations away from the spheroidal and the occurrence of straight-edged images are stressed again.

INTRODUCTION

NO experiments have been reported concerning the shadow-casting of carbon black, probably because the diameters and surface details of the ultimate units are obscured somewhat when the metallic deposits are too heavily applied. However, if the metal is kept reasonably light, useful morphological detail is obtained without loss of accurate particle size data. Shadow-casting techniques were used by us in certain experiments on the aggregational properties of carbon blacks in order to determine the heights of the aggregates,* and from these experiments a number of incidental observations were made which were sufficiently fundamental of carbon black to be reported separately.

The number of micrographs which can be shown to illustrate each observation is limited by space but those which are shown are representative, and the effects are reproducible for the samples. Although some of them may seem startling at first examination, the morphological characteristics reported are consistent with existing knowledge and theories of carbon black formation.

It is not the purpose of this discussion to suggest that all carbon blacks illustrate the characteristics described, or that all particles of one black do so, but that such isolated particles or groups can and do occur.

Since carbon black aggregates are relatively large and high, a rather steep angle of shadowing is necessary, and in general, it was kept considerably greater than 45 degrees. The angle of shadow-casting has a tangent of 1.6 for all of the micrographs discussed. Chromium was used in all cases. The micrographs were taken using an objective aperture of 40 microns.

RETICULATE CHAIN STRUCTURE

Ladd and Wiegand,[1] using the terminology "reticulate chain structure," have claimed from electron micrographs of unshadowed specimens that in some carbon blacks the individual particles or units are joined one to the other by short necks or rods. The observations reported here have shown nothing to discredit this claim but have provided further indications that such necks can and do exist in carbon black. Their existence is readily demonstrated in Thermax,** a medium thermal rubber reinforcement black with mean diameter 3200 A.U. They are observed frequently, too, in P-33,** a fine thermal black with mean diameter 1700 A.U., also used in rubber reinforcement. Presumably, they should exist in other blacks of even lower mean particle sizes, but obviously would be less easily demonstrable. It is known that many of the aggregates in carbon black do go through rigorous industrial processes (e.g., rubber compounding) without breaking up. From this, one may conclude that the units in such aggregates are joined rigidly together and by inference, that some sort of joining necks exist between the particles making up these aggregates.

Some of the micrographs are reproduced twice because of the different significant observations to be made from those which are over and those which are under developed in the photographic printing processes. The former give most clearly the information which is obtainable from the shadow; the latter that which is most readily obtained from observing the particles themselves. Thus, the parts of Fig. 1 are each reproduced in lighter tones in Fig. 2, and those in Figs. 3A, B, C, D, and E in Figs. 3F, G, H, I, and J, respectively.

The rods referred to by Ladd are visible in Figs. 2A and B and have lengths of a little more than 135 A.U. in Fig. 2A, and about 240 A.U. in Fig. 2B. The rods in Figs. 2D and E are also about 135 A.U. long. Figures 1 and 2C are reproduced to show that the effect is not due to lack of resolution, because in these figures two particles apart by less than 135 A.U. are clearly separated, and yet no neck is demonstrable, similarly in Figs. 1 and 2E. Lack of resolution probably accounts for the fuzziness between the particles in Fig. 1 and 2F, but these particles are separated by less than 55 A.U., which is well below the dimensions under consideration. From this last observation, one may conclude that the resolution in these micrographs is better than 55 A.U.

It may be argued that another particle either over or under the group and oriented between two particles

* Unpublished work by L. H. Cohan and J. H. L. Watson.

[1] W. A. Ladd and W. B. Wiegand, *The Rubber Age*, June 1945.

** Thermatomic Carbon Company, Sterlington, Louisiana.

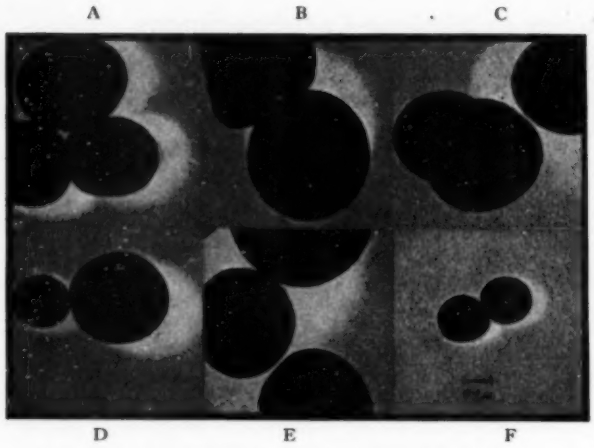


FIG. 1. Electron micrographs of Thermax carbon black. Heavy development. ×37,500.

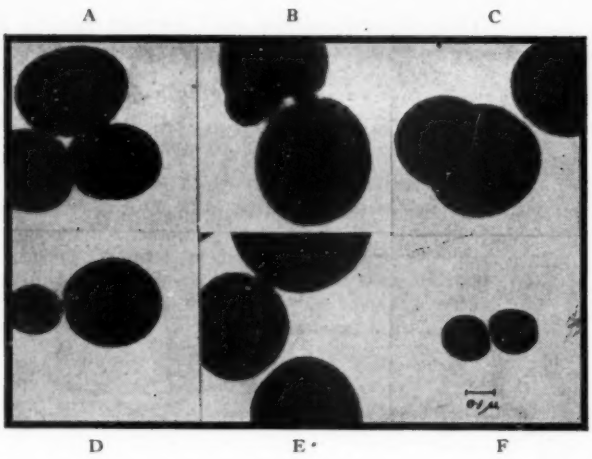


FIG. 2. Electron micrographs of Thermax carbon black. Same fields as Fig. 1; light development. ×37,500.

might account for the rods. If so, the appearance of the micrographs in the region of the rods in Fig. 2 should be similar to those in Fig. 3, but they are not. In addition, the necks do not throw shadows as particles would be expected to do.

In Figs. 3A, B, and C particles are illustrated which are touching but which are not thought to be joined rigorously together. Figure 3F shows considerable overlapping in which the whole outline of the particles is easily visible. In Fig. 3G another overlapping pair is shown where the overlapping is not quite as extensive but where again the whole outline is visible. Even in Fig. 3H where the overlapping area is very small the outline is still visible and has not the same appearance as that which is interpreted as a neck in Fig. 2.

That particles grow intimately together is shown further in Figs. 3D and 3I where twin particles are seen. There is no overlapping, but two particles have apparently stuck together in the flame and grown when additional carbon collected upon them. Both sides of the doublet throw the same length of shadow, proving that one is not sitting upon or under the other.

Most of the particles discussed above were taken from a Thermax sample and were studied under identical conditions as a comparison. Figures 3E and 3J show an example of a P-33 doublet which is joined by a strong neck about 100 A.U. long.

It is not always possible to find two particles so oriented that they are simultaneously in the plane of the supporting film and perpendicular to the direction taken by the shadowing metal. However, some cases of this have been observed, and one is reproduced in Figs. 4A and B. If two separate particles which are just touching are involved here, the shadows might be expected to proceed as the dotted lines in the micrograph, and if an overlapping particle were involved its outline should be visible, but since neither observation is made in the actual micrographs, one can conclude that the two particles are joined together by a short neck about 135 A.U. long and with a diameter of 1400 A.U. In addition, it is noted in this particular case that the neck enters one of these particles in such a way that it throws a shadow upon the particle itself. This phenomenon is also visible in Fig. 2D where a rod is seen leaving one particle and "dimpling" into another. Dumb-bell shaped shadows have not been observed because of insufficient resolution, lack of shadow sharpness and the short length of the necks involved.

From these studies it is concluded that the observations reported by Ladd are supported by shadow-casting data, and that rigid aggregates of carbon particles exist in Thermax and P-33. Since the necks are exceedingly short as compared with particle diameter (see Table I), it is probable that such necks would not be resolved in any but the larger sized blacks, but their existence in the smaller sized carbons is indicated indirectly from rubber data.

The average length of neck for Thermax is 160 A.U. and the mean diameter of Thermax is 3200 A.U., giving a percentage ratio for $l:d$ of 5 percent. If a similar ratio of neck length to mean particle diameter is assumed to exist for other blacks, the average neck length would be but 85 A.U. in P-33. In any but the best micrographs these would not be readily seen, and consequently the number of necks observed in P-33 would be less than in Thermax. In carbon blacks with mean sizes less than P-33 (and these constitute the great majority of them) the necks could not be resolved at all. This probably explains why they are observed almost exclusively in Thermax, only slightly in P-33 and not at all in the others.

PARTICLE DENSITY

Particles which appear hollow in electron micrographs have been observed in shadowed specimens of P-33 and Thermax carbon blacks. Several examples are shown in the micrographs of P-33 in Fig. 5. In this figure examples have been chosen for a large range of particle size, from 420 A.U. in 5B to 3250 A.U. in 5F, in order

to show that the hollow particles can occur throughout a wide range of particle sizes. However, their mean diameters and the major portion of their distributions are considerably less in size than they are for the whole distribution. It will be seen that the hollow units occur in groups, doublets or singlets just as do the more usual particles of carbon black.

Possible explanations for the appearance of hollowness might be that (1) the particles are actually hollow, (2) the particles have interior regions which are less dense than the peripheral areas, or (3) the particles have a relatively deep surface depression. Since there is such a difference of intensity between the central and

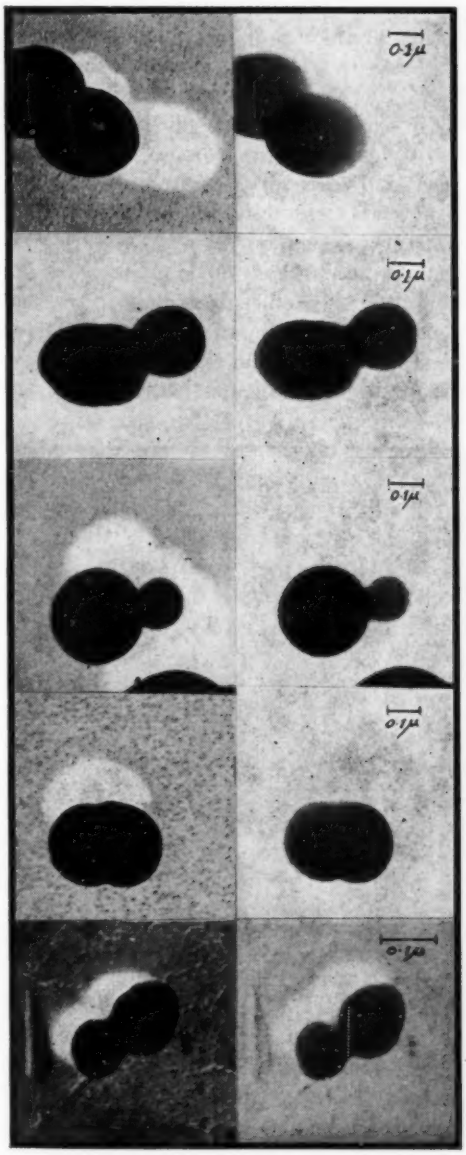


FIG. 3. Electron micrographs of carbon black. A, B, C, F, G, H of Thermax, to show the appearance of various degrees of particle overlapping; D and I of Thermax, ×45,000, to show a rigid aggregate, no overlapping. E and J, ×72,000, to show a rigid aggregate of two particles in P-33.

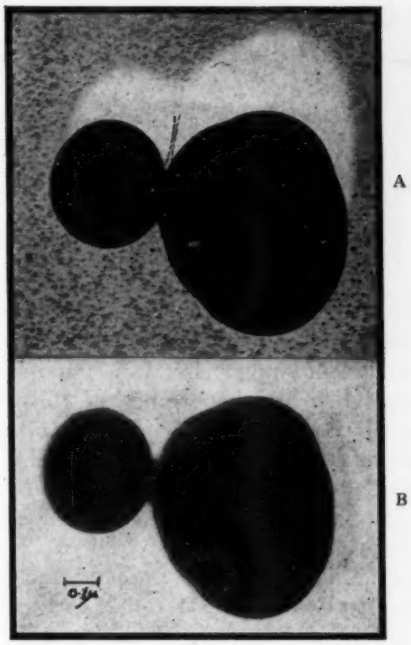


FIG. 4. Electron micrographs of Thermax carbon black. (A) heavy development; (B) light development. ×45,000.

the outer regions of the particle images, it would seem that at least some of the particles must be describable under class 1. However, there is much to suggest that the "hollowness" is merely lack of carbon density with perhaps the occasional occurrence of a vacuole or granule in the central regions. This is indicated by the variations in density which are seen over the particle images and which cannot be interpreted as surface roughness, since they do not cast shadows. In Fig. 5 examples have been chosen which show little variation in density in the internal regions, and which consequently might belong to class 1 rather than 2. In Figs. 5B and C, however, doublet particles are seen in which a common dense, dividing area exists between the two units in the particles. This would indicate that two separate particles stick together in the flame during the formation of such doublets and that additional carbon deposits upon them to join them rigidly together into a unit. Shadow-casting raises the contrast and shows up the division more clearly. Since the tendency of P-33 is to exist chiefly as singlets or doublets, the large group in Fig. 5A is probably a loose aggregation only.

Although case 3 might explain the appearance of the shadow-cast specimens, it is extremely doubtful that such particle shapes exist, and in addition, when lying on their side they should appear cylindrical. No cylindrical images are seen. Also, judging from the opaqueness of usual carbon particles to electrons, the depressions would have to be very deep, and the wall high to give the variations in intensity observed. Such morphological extremes should be visible in unshadowed samples but

TABLE I. Data for length of neck and particle diameter for several examples in Thermax.

l length of neck (A.U.)	d particle diameter (A.U.)	$\frac{l}{d}\times 100$
135	3740	3.6
135	3740	3.6
240	4400	5.4
160	4400	3.6
135	3200	4.2
135	3600	3.8

they are not seen. Figure 9A shows an opaque and a non-opaque particle of similar size. It is impossible to interpret absolutely the appearance of this particular non-opaque particle. It could be either, one with internal volume less dense than the periphery, or one with a depressed area in its surface. This, however, is the only example observed which had definite indication of such a depression.

The explanation for the effect appears most logically to be a combination of cases 1 and 2. In Fig. 6 several examples are given of particles whose density is variable. In these examples this does not seem to be due to surface morphology but to an internal lack of homogeneity in the carbon particle. Each of the three examples of P-33 in Fig. 6 is reproduced with both light and heavy development of the prints in order to give clearly both the shadow and the internal structure. Figures 6A and B show a doublet wherein the hollowness is continuous from particle to particle through the neck. A piece of opaque carbon of comparable size is also seen in the same figure. In 6C and D a particle of non-uniform density is seen. Figures 6E and F show one particle with a dense area in the center which is supported by strands of a dense material reaching to the sides to divide the particle approximately into quadrants. The second particle in Figs. 6E and F is a triplet with one particle easily differentiated from the first, but with another apparently encased within a larger one.

It is concluded that some particles of carbon black have complex sub-structures (not to be confused with the crystallite structure), and from these observations and others it is concluded further that this structure may take many variations. The sub-structures within the particles are not thought to be pseudostructures[2] introduced by the electron beam because no effect other than growth[3] due to contamination has ever been observed in carbon black, regardless of intensity or duration of electron bombardment. In addition, both hollow and solid particles of comparable sizes are found in identical fields treated in identical ways in both P-33 and in Thermax.

In specimen preparation carbon alone was blown dry upon the supporting films. The hollow particles have all the characteristics of size, shape and distribution which are expected normally of carbon blacks in electron micrographs except for the new observation of

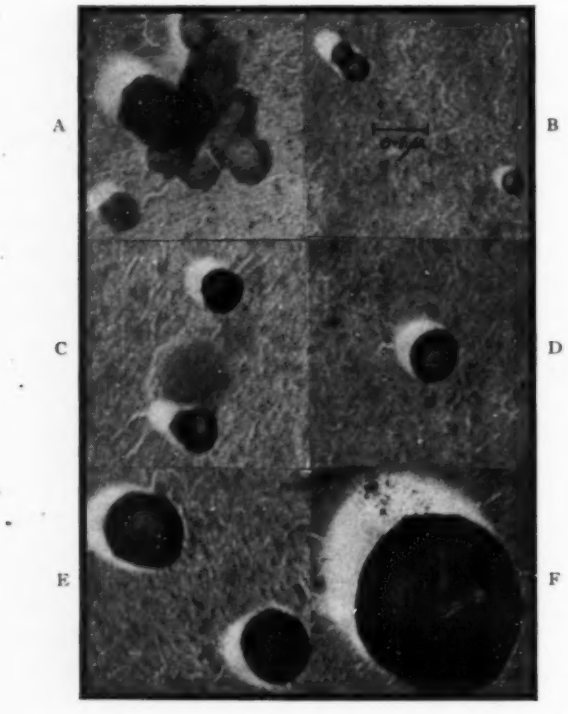


FIG. 5. Electron micrographs of hollow particles in P-33 carbon black. ×72,000.

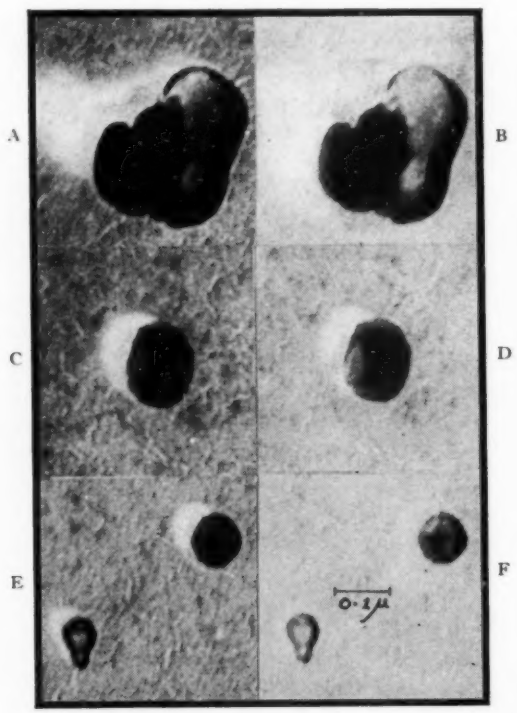


FIG. 6. Electron micrographs showing internal inhomogeneities in hollow particles of P-33 carbon black. ×72,000.

[2] John H. L. Watson, J. App. Phys. **19**, 713 (1948).
[3] John H. L. Watson, J. App. Phys. **18**, 153 (1947).

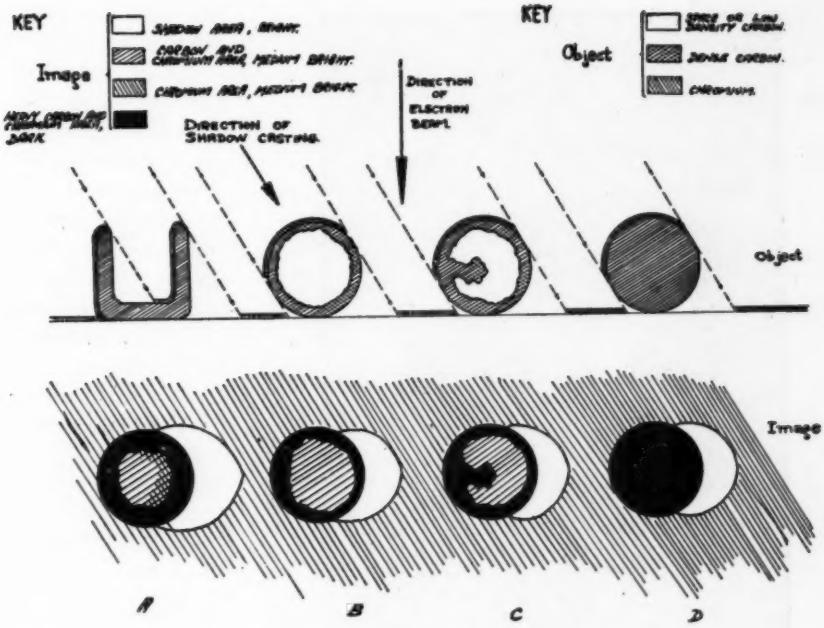


FIG. 7. Schematic drawing to show possible mechanisms for image formation of hollow particles of carbon black.

intensity gradation over the images. It is true that most of the hollow particles are small, but they are stil within ranges which are often encountered.

Hollow particles have not been observed in unshadowed specimens. However, it is doubtful that the effect can be an artifact introduced by the shadowing technique, because the dark image areas (thick or dense regions in the object) proceed completely around the periphery of the image, and neither the angle of shadowing nor the amount of deposited chromium are such as to cause a "collar" to form on the particles. In addition, both opaque and non-opaque particles, of comparable size, are observed together in the same shadowed fields. The shadowing has apparently raised the contrast sufficiently to make the effect visible.

Possible mechanisms of the image formation are illustrated in Fig. 7. Figure 7A represents the image appearance which might be expected from a particle with a depressed surface. Actually, no image had intensity gradations to agree completely with this picture. On the other hand, the images in Figs. 7B, C and D are quite common in the micrographs and can be explained through considerations of the effectively greater scattering power of the peripheral regions of the particles projected into the image plane, this scattering power being enhanced by the shadowing process.

It is possible to secure an artifact of hollowness in particles of low density and low atomic weight if a sample is shadow-cast too heavily at an acute angle, especially, if the photographic plate is over-exposed and developed in a contrast developer. This set of conditions has been applied to give the peculiar appearance in Fig. 8, which is an electron micrograph of a chromium shadow-cast emulsion prepared according to the conditions mentioned above. The particles appear to be hollow in this figure, although from other work it is known that they are not. However, in the case of the P-33 and Thermax the shadow-casting was done at the relatively large angle of 58 degrees, and neither the amount of deposited chromium, nor the plate exposure was exorbitant. In addition, a fraction of the particles are hollow in the blacks, whereas in the emulsion all of them appear so.

The fact that particle sizes in the distribution of the hollow units are less[4] than those of the whole sample,

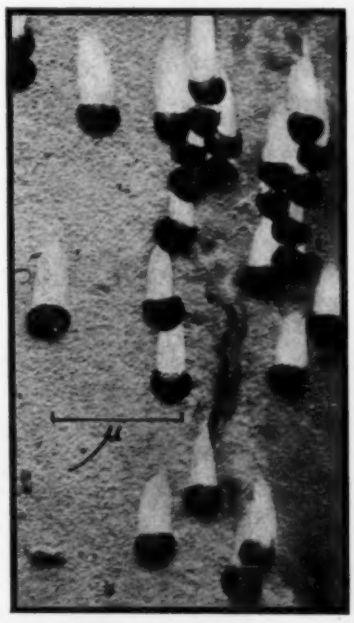


FIG. 8. Electron micrograph of a polymer emulsion in which an artifact of "hollowness" is achieved by shadowing heavily at about 10 degrees and overdeveloping the negative. ×17,250.

[4] John H. L. Watson, The Electrochemical Society, Preprint 92-4 (1947).

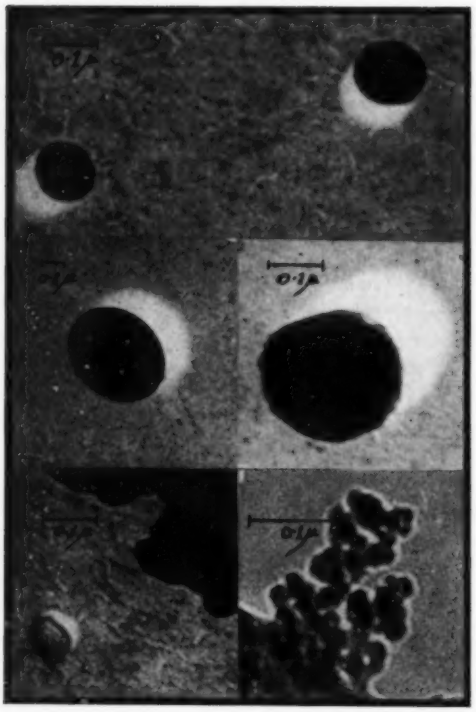


FIG. 9. Electron micrographs of carbon blacks: (A) P-33 (×72,000), (B) Thermax (×30,000), (C) P-33 (×72,000), (D) P-33 (×72,000), (E) Witco Super (×120,000).

coupled with the structural appearance of some of the particles, Fig. 9C and D, might suggest that this group is a crystalline component of P-33. Crystals of various shapes have been reported in other carbon blacks.[4,5] Figure 9D illustrates that crystals can exist in P-33, but such definite crystals as this are observed very infrequently in it. Where the intensity over the image of the particle in Fig. 9D may be explained by assumptions of crystallinity for this particle, it is not probable that this can account for the hollow appearance of the others, because the vast majority of them are rounded and non-crystalline in appearance and because no change in the image is noted with variations in beam aperture. However, this explanation for the hollowness must be considered as an alternative, especially as intensity variations over images of crystalline material have been reported[6] earlier.

In the smallest particles, such as those of Witco Super (mean diameter about 80 A.U.) the effects of Fresnel diffraction in lack of focus are often such as to give an appearance of hollowness, Fig. 9E, but this cannot explain the hollow appearance of the *large* particles in P-33 and Thermax.

PARTICLE SHAPE AND SURFACE MORPHOLOGY

The study of the surface of carbon black particles is important because of the effect which surface properties have on such things as rubber reinforcement, covering power, color, etc. It is also important because of the information it may yield concerning the fundamental study of the mechanism of carbon black formation.

Figure 9B is shown to emphasize the elliptic rather than the usually reported circular nature of the electron images of normal carbon particles. The particle in this figure has a major diameter of 1920 and a minor diameter of 1590 A.U., giving a ratio of 1.21. Its height is approximately 1470 A.U. This ellipticity, which is found to some extent in every carbon sample studied, is suggested as another important measurable property. From examination of micrographs of a large representa-

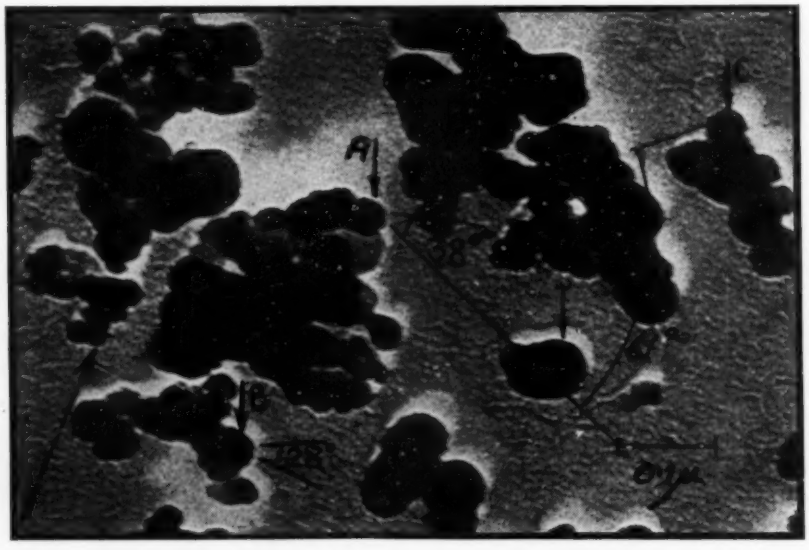


FIG. 10. Electron micrograph of Sterling 105; at the arrows are particles which exhibit a laminar structure. ×78,000.

[5] John H. L. Watson, J. App. Phys. **17**, 121 (1946).
[6] R. D. Heidenreich and L. Sturkey, J. App. Phys. **16**, 97 (1945).

tion of carbon blacks, the author is convinced that divergence from the spherical is the rule rather than the exception for carbon particles. Quantitative studies have been made, and statistical averages of major to minor axes have been measured and calculated. For a series of blacks*** studied recently such values are: Continental A, 1.26; Continental AA, 1.21; Continex FF, 1.26; Continex HMF, 1.16 and Continex SRF, 1.07.

Figure 9C indicates that it is possible in light field to resolve straight edges which could yield the Bragg reflections which are seen in dark field.[7] The particle in Fig. 9C is of P-33. It seems to be made up of a flaky material with appearance reminiscent of graphite in electron micrographs. There are a number of straight edges around its circumference with dimensions of 1250, 540, 750, 420, 540 and 830 A.U. The hexagonal unit of P-33 in Fig. 9D referred to in the preceding section has straight edge dimensions of 500 and 250 A.U.

Intensity differences in electron images, especially of shadow-cast specimens, may also be a function of surface morphology. In Fig. 10 is shown a shadowed sample of Sterling 105,† which is a reinforcing furnace black with mean diameter about 360 A.U. At the arrows are particles possessing a laminar structure which appears to extend throughout the whole volume. The dark areas, which represent heights are in straight lines but are observed within some particles to change direction sharply. The particle at *A* is 300 A.U. high and has three laminar lines over its surface, each separated by a distance of about 100 A.U. (all values are quoted to better than 10 percent). The same distances are obtained from measurements made on other particles which demonstrate the laminations in this micrograph. Since particles separated by much less than 100 A.U. are resolved, it is concluded that the resolution is sufficient to show the structures without artifact and that they are real.

Serrations which conform in orientation and in dimensions with the laminar structures are noted along the edges of particles. In unshadowed specimens roughened particle edges are also observed in micrographs of Sterling 105 and in other blacks, but neither the definiteness of the serrations nor the laminations themselves have been observed in unshadowed samples.

The majority of the serrated edges are roughly parallel to the direction of shadowing, and most of the laminations are oriented nearly perpendicular to this direction. Both of these facts are to be expected if a real phenomenon is being demonstrated by virtue of the extra contrast the shadowing affords. One would also expect that the laminations and the serrations would still be visible within a certain range of angle on either side of the direction perpendicular or parallel, respectively, to the shadow. This is found to be the case in the micrographs where from actual measurements serrated edges and laminar lines are found to be observable up to about 52° on *either* side of the optimum positions.

TABLE II.

Carbon black sample	Mean particle diameter (A.U.)	Image description* Striated particle surface	Serrated particle edges	Rough particle surface
Carbolac 1	90	Suggested but unresolved		
Continental A	265	*M*	*M*	*M*
Continental AA	330	*M*	*F*	*M*
Sterling 105	360	*M*	*M*	*M*
Continex FF	430	*F*	*M*	*N*
Continex HMF	600	*F*	*M*	*F*
Continex SRF	980	*N*	*F*	*F*
P-33	1700	*N*	*F*	*N*
Thermax	3200	*N*	*N*	*N*

* *M*—many occurrences; *F*—few occurrences; *N*—no occurrences.

Several representative angles are noted on Fig. 10. Within particle *A* the laminar lines bend at an angle of about 38 degrees, and in the surface of particle *B* at an angle of about 28 degrees. The lines in particle *C* are at an angle of 61 degrees to those in particle *A*. The approximate direction of shadowing is given by the broken arrow in the lower left-hand corner of the figure.

As is usual in shadow-cast preparations of film-supported specimens, there is background structure. This is not observed in the shadows and presumably cannot be visible in the protected areas under the particles. Therefore, it is not this structure which is responsible for the laminar structures seen in the particles. Neither can it account for the serrated edges, since these edges are visible in the shadows.

The proportion of particles showing these structures in Sterling 105 is so high that it is unlikely that their presence is a chance affair. For example, in the small field shown in Fig. 10 there are at least nine particles with striations over their surface. Other black samples have been examined after shadow-casting, and in many cases somewhat similar structures have been observed. This has been true particularly of carbon blacks of the intermediate sizes (mean diameter about 500 A.U.).

Close examination of the plates discloses that the laminar lines themselves are discontinuous with dark areas which are about 50 A.U. in diameter or less repeating along their lengths. Such dark areas which may represent small "bumps" or "blisters" on the surfaces of the particles are observed in many of the shadow-cast blacks which have been examined, and indicate again that the particle surfaces may be rough rather than smooth. It has been impossible so far to reproduce these intensity variations in prints at a size and contrast suitable for publication, but the structures being discussed in this paper are all larger than the grain size of the photographic emulsion in the prints and plates.

The rough appearance is present in most of the shadowed specimens except P-33 and Thermax where the image intensity is homogeneous over the particle

*** These samples were supplied by the Witco Chemical Company, 295 Madison Avenue, New York, N. Y.

[7] C. E. Hall, J. App. Phys. 19, 271 (1948).

† Godfrey L. Cabot, Inc., Boston, Massachusetts.

surface, with the exception of those cases reported in the third section of this paper. In P-33, however, serrated edges are often observed. A tabulation of the structures occurring in a number of samples is given in Table II. It will be observed that as the particle size increases less structure is seen. This suggests that the amount of particle surface structure and roughness decreases with increased particle size.

SUMMARY

The following observations of structure and shape in shadow-cast carbon black specimens have been made: (1) necks joining particle to particle have been demonstrated, (2) particles which are less dense inside than toward the periphery and substructures within such "hollow" particles have been discovered, (3) striated surfaces strongly suggestive of a laminar structure have been seen in some particles, (4) serrated edges which conform in orientation with the striations have been detected, (5) rough areas arranged either in lines or at random are observed over the surfaces of some particles, (6) shape variations away from the spheroidal and toward the prolate spheroidal or ellipsoidal have been mentioned, along with (7) the occurrence of straight-edged images (indicative of flat planes on particle surfaces).

Coexistence Relations of $n+1$ Phases in n-Component Systems

PAUL M. PEPPER

Department of Mathematics, University of Notre Dame, Notre Dame, Indiana

(Received December 14, 1948)

In an investigation of the coexistence relationships of $n+1$ phases in the isothermic-isobaric sections of n-component systems, Paul A. Beck found that there are at most two coexistence patterns possible. Furthermore, theoretically, only a well-chosen sample need be examined for the phases it contains in order to clarify the phase coexistence relationships of the $n+1$ phases whose compositions are already known. The present paper gives formulas for calculating the composition of the "well-chosen sample." Supplementary formulas are given for the range of coexistence, in the composition space, for any n phases known to coexist.

INTRODUCTION

IT was shown by Beck[1] that in an n-component system for any $n+1$ phases of known composition, independent in sets of n, there is a unique sample composition, P_Z, of such a nature that experimental determination of the phases coexisting in P_Z at a certain temperature and pressure fully clarifies the coexistence pattern of the $n+1$ phases under the conditions considered. Thus, if P_Z is known, theoretically a single experiment is sufficient to determine the coexistence relations of $n+1$ phases. It is, therefore, of great practical interest to develop methods of calculating the actual composition of the sample P_Z in a given system. The purpose of the present paper is to develop such a method. Vector methods are used, due to their adaptability to the problem.

When the phase coexistence pattern is determined (it is known which phases can coexist and which ones cannot), the composition space may be divided into n-phase fields (composition ranges within each of which the same n phases coexist), provided that the composition ranges of all of the single phase fields are narrow. In such a case, each phase may be reasonably well represented by a single composition; the hyperplanes which separate the n-phase fields from each other may be calculated from the compositions of the phases. A further objective of the present paper is to describe a general mathematical procedure which leads to a system of linear inequalities delineating any one of the n-phase fields as an n-hedron in the composition space. With that information it is possible to calculate for a sample of any composition within the range of the $n+1$ phases (in equilibria at the temperature and pressure considered) which phases it will consist of.

The author wishes to express his appreciation to Paul A. Beck, who suggested the problem and assisted with the metallurgical terminology.

FORMULAS FOR THE COMPOSITION P_Z

In an n-component system the components involved are to be denoted by $e_1, \cdots, e_n$ and each phase P is thought of as a vector in *n-dimensional space*. If the vector is described by noting the amount of each component entering into the phase, the coordinates will be called $a_1, \cdots, a_n$ and the vector will be denoted by A; that is, a_j is the amount of the component e_j which enters into P. Clearly, any numbers proportional to $a_1, \cdots, a_n$ suffice to determine the phase, so that the vector cA with coordinates $ca_1, \cdots, ca_n$ ($c \neq 0$) will be an equally good representation of the phase P. The particular system, where numbers of atoms are the coordinates of the vectors, will be called the *atomic* coordinate system, and the vectors, *atomic vectors* representing P. Due to the simplicity of numerical computations in this coordinate system, it will be used wherever possible.[2] On the other hand, the method is

[1] Paul A. Beck, J. App. Phys. **16**, 808–815 (1945).

[2] If the components are chemical compounds, the numbers of molecules and the molecular weights may be substituted for the

general and formulas are given throughout for the use of weights or weight fractions. For example, see Eqs. (1′), (2′), and (4′) of this section.

Let $P_1, \cdots, P_{n+1}$ be $n+1$ phases in an n-component system, each n of which are independent. That is, no n of the associated vectors lie in an $(n-1)$-dimensional space. Let $A_1, \cdots, A_{n+1}$ be the atomic vectors representing them. Let $a_{i1}, \cdots, a_{in}$ be the coordinates of A_i, so that a_{ij} is the number of atoms (molecules) of the jth chemical element (component) in the ith phase. Let ω_j be the atomic (molecular) weight of the jth element (component).[3]

When the compositions of the $n+1$ phases $P_1, \cdots, P_{n+1}$ are given, the numbers a_{ij} are known. As the first step in determining the composition P_Z, solve the vector equation

$$\alpha_1 A_1 + \cdots + \alpha_n A_n + \alpha_{n+1} A_{n+1} = 0, \text{ with } \alpha_{n+1} = -1, \quad (1_v)$$

or, what is the same, the system of linear equations

$$\begin{aligned} &\alpha_1 a_{11} + \cdots + \alpha_n a_{n1} + \alpha_{n+1} a_{n+1,1} = 0, \\ &\quad \cdot \\ &\quad \cdot \\ &\quad \cdot \\ &\alpha_1 a_{1n} + \cdots + \alpha_n a_{nn} + \alpha_{n+1} a_{n+1,n} = 0, \\ &\qquad\qquad\qquad\qquad \alpha_{n+1} = -1, \end{aligned} \quad (1)$$

for the unknowns $\alpha_1, \cdots, \alpha_n, \alpha_{n+1}$. An interpretation of the numbers α_i is given following Eq. (2). If A_Z, with coordinates $a_1, \cdots, a_n$, is the atomic vector representing P_Z, then

$$A_Z = \sum_{\alpha_i > 0} \alpha_i A_i = -\sum_{\alpha_i < 0} \alpha_i A_i, \quad (2)$$

the first summation being taken over values of i for which $\alpha_i > 0$ and the second over those values of i for which $\alpha_i < 0$. The coefficients α_i and $-\alpha_i$ are then parameters representing the amounts of the various phases which may be used to compose P_Z. (See also Appendix I.) Both vector sums in (2) are to be computed in determining A_Z. Their equality checks the solution of Eqs. (1) as well as the computation of the coordinates of A_Z from the A_i and α_i. Any set of numbers proportional to $a_1, \cdots, a_n$ represents the phase P_Z, thus any common factor of the coordinates of A_Z may be discarded.

The result of Beck becomes: Either, all those phases with positive α_i may coexist, or all those with negative α_i, but not both. Which set may coexist is to be determined by experiment.

Once the number of atoms of each chemical element in P_Z is determined, a conversion to weight fractions

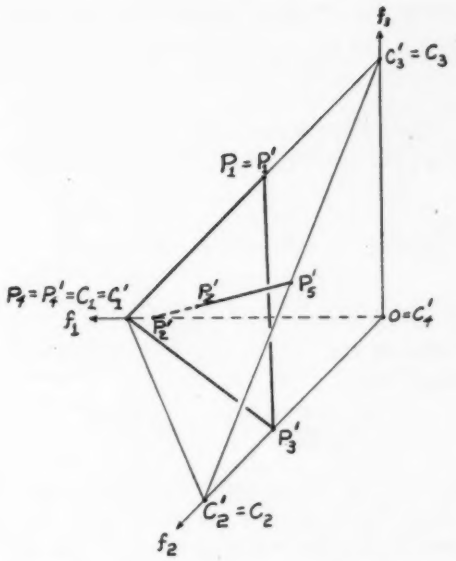


FIG. 1. The separation, for the example of the text, of the five phases into the two uniquely determined groups, as described by Beck, is shown here. The two groups are P_1, P_2, P_4 and P_2, P_5; the segment connecting P_2 to P_5 pierces the triangle determined by P_1, P_3 and P_4 at P_Z. Experiment, using the composition P_Z, shows which group comprises the coexisting phases.

may be made in the following way. Let

$$w_j = a_j \omega_j, \quad j = 1, \cdots, n, \quad (3)$$

and define

$$w = \sum_{j=1}^{n} w_j, \quad f_i = w_i / w, \quad i = 1, \cdots, n. \quad (4)$$

Then $f_1, \cdots, f_n$ are the weight fractions of the respective chemical elements as they occur in P_Z.

A check on the numerical work done to obtain the numbers w_j may be made as follows: Let $w_{ij} = a_{ij}\omega_j$, $(i = 1, \cdots, n+1;\ j = 1, \cdots, n)$ and let W_i be the vector with coordinates $w_{i1}, \cdots, w_{in}$. Let W_Z denote the representation of P_Z in this system, then the coordinates $w_1, \cdots, w_n$ of W_Z (the same coordinates as computed in (3)) may also be computed from the vector relation

$$W_Z = \sum_{\alpha_i > 0} \alpha_i W_i = -\sum_{\alpha_i < 0} \alpha_i W_i \quad (5)$$

where the α_i are those computed from Eqs. (1).

The set of weight fractions corresponding to P_Z will be regarded as the coordinates of a vector F_Z. A check on the computation of the coordinates of F_Z is made in the usual way by means of the relation

$$f_1 + \cdots + f_n = 1. \quad (6)$$

In case the components are not chemical elements (or, if desired, even if they are elements) the coefficients a_{ij} of Eqs. (1) may be replaced by coefficients u_{ij}, where u_{ij} represents the weight of the jth component in the ith phase. In this case we will denote the

number of atoms and the atomic weights. Such a coordinate system will be called a *molecular* coordinate system and the vectors, *molecular vectors*.

[3] Thus if e_3 is tin (Sn) and P_4 is a tin compound in which each molecule has two atoms of tin, then in the atomic coordinate system $a_{43} = 2$ and $\omega_3 = 118.7$.

unknowns by $\beta_1, \cdots, \beta_n, \beta_{n+1}$ and write the equations as

$$\beta_1 U_1' + \cdots + \beta_n U_n + \beta_{n+1} U_{n+1} = 0, \quad \beta_{n+1} = -1 \quad (1_v')$$

or

$$\begin{aligned} &\beta_1 u_{11} + \cdots + \beta_n u_{n1} + \beta_{n+1} u_{n+1,1} = 0, \\ &\quad \cdot \\ &\quad \cdot \\ &\quad \cdot \\ &\beta_1 u_{1n} + \cdots + \beta_n u_{nn} + \beta_{n+1} u_{n+1,n} = 0, \\ &\qquad\qquad\qquad\qquad \beta_{n+1} = -1, \end{aligned} \quad (1')$$

where U_i is the vector with coordinates $u_{i1}, \cdots, u_{in}$.

The solution $\beta_1, \cdots, \beta_n, \beta_{n+1}$ so obtained is to be substituted into

$$U_Z = \sum_{\beta_i > 0} \beta_i U_i = -\sum_{\beta_i < 0} \beta_i U_i \quad (2')$$

to obtain U_Z, the representative of P_Z in that coordinate system. Since the u_{ij} are weights of the various components in P_i, the coordinates $u_1, \cdots, u_n$ of U_Z are likewise weights. It suffices to substitute the coordinates of U_Z into

$$u = \sum_{j=1}^{n} u_j, \quad f_i = u_i/u, \quad i = 1, \cdots, n \quad (4')$$

to obtain $f_1, \cdots, f_n$, the weight fraction coordinates of P_Z. These weight fraction coordinates are the fractional parts of the *components* as they enter into P_Z. As a precaution against misuse of the formulas, it is recommended that the coordinates $u_{i1}, \cdots, u_{in}$ be weight fractions or weight percentages of the respective components in the ith phase, although this is not necessary with correct use of Eqs. (1′) and (2′).

We compute a hypothetical example to illustrate the first method. Let a four component system be given in which the chemical elements $e_1, \cdots, e_4$ are the components. Let the atomic weights of the respective elements be $\omega_1 = 49.3$, $\omega_2 = 41.5$, $\omega_3 = 29.7$ and $\omega_4 = 17.3$. Let the atomic vectors representing the five phases be

$$A_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix}, A_2 = \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \end{bmatrix}, A_3 = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 3 \end{bmatrix}, A_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, A_5 = \begin{bmatrix} 0 \\ 8 \\ 9 \\ 3 \end{bmatrix},$$

where the coordinates of the vectors are written in vertical columns since this is the way they appear as coefficients in the Eqs. (1). For the hypothetical example the Eqs. (1) read

$$\begin{aligned} \alpha_1 + 3\alpha_2 \qquad + \alpha_4 \qquad &= 0, \\ 2\alpha_3 \qquad + 8\alpha_5 &= 0, \\ 2\alpha_1 \qquad\qquad + 9\alpha_5 &= 0, \\ \alpha_2 + 3\alpha_3 \qquad + 3\alpha_5 &= 0, \\ \alpha_5 &= -1. \end{aligned}$$

The solution of these equations is $\alpha_1 = 4.5$, $\alpha_2 = -9$, $\alpha_3 = 4$, $\alpha_4 = 22.5$ and $\alpha_5 = -1$. From Eq. (2) it follows that $A_Z = 4.5A_1 + 4A_3 + 22.5A_4$ or $A_Z = 9A_2 + A_5$, so that

$$A_Z = \begin{bmatrix} 27 \\ 8 \\ 9 \\ 12 \end{bmatrix}.$$

Thus in P_Z there are 27 atoms of e_1, 8 atoms of e_2, 9 atoms of e_3 and 12 atoms of e_4, or numbers proportional to these.

The result of Beck implies that either P_2 and P_5 (with negative α_i) coexist or P_1, P_3 and P_4 (with positive α_i) coexist. If experiment shows that P_2 and P_5 coexist, then P_1, P_2, P_3, P_5; P_1, P_2, P_4, P_5; P_2, P_3, P_4, P_5 are the sets of phases which can coexist. If experiment shows P_1, P_3, P_4 coexist, then the sets P_1, P_2, P_3, P_4 and P_1, P_3, P_4, P_5 can coexist. See Fig. 1 for a projection of the 4-dimensional diagram into the $f_1 f_2 f_3$-space.

Conversion of the coordinates of A_Z to weights by Eqs. (3) gives

$$w_1 = 1331.1, \quad w_2 = 332.0, \quad w_3 = 267.3, \quad w_4 = 207.6$$

as coordinates of the vector W_Z. Conversion to weight fractions by Eqs. (4) yields

$$f_1 = 0.624, \quad f_2 = 0.155, \quad f_3 = 0.125, \quad f_4 = 0.097$$

as weight fractions of the respective elements. The desired sample, P_Z, is then composed of 62.4 percent of e_1, 15.5 percent of e_2, 12.5 percent of e_3 and 9.7 percent of e_4 by weight. A check by Eq. (6), of the conversion from weights to weight fractions shows this within the allowable error.

The numbers w_i should also be checked, however, and for this, first compute the special vectors, W_i, representing the phases $P_1, \cdots, P_5$ in the weight system,

$$W_1 = \begin{bmatrix} 49.3 \\ 0.0 \\ 59.4 \\ 0.0 \end{bmatrix}, \quad W_2 = \begin{bmatrix} 147.9 \\ 0.0 \\ 0.0 \\ 17.3 \end{bmatrix}, \quad W_3 = \begin{bmatrix} 0.0 \\ 83.0 \\ 0.0 \\ 51.9 \end{bmatrix},$$

$$W_4 = \begin{bmatrix} 49.3 \\ 0.0 \\ 0.0 \\ 0.0 \end{bmatrix}, \quad W_5 = \begin{bmatrix} 0.0 \\ 332.0 \\ 267.3 \\ 51.9 \end{bmatrix},$$

where $w_{ij} = a_{ij}\omega_j$. When these are combined vectorially by Eqs. (5), using the α_i determined by Eqs. (1), the same coordinates of W_Z are found as before.

As an illustration of the second method, we compute another hypothetical example. Suppose the weight fractions or weight percentages of the various components in the given phases are known. These may be regarded as weights and, as such may be taken as coefficients in the Eqs. (1′). For this example the weight percentages of the components in P_1 are assumed to be 45.4, 0.0, 54.6 and 0.0; in P_2 to be 89.5, 0.0, 0.0, 10.5; in P_3 to be 0.0, 61.6, 0.0, 38.4; in P_4 to be 100.0, 0.0, 0.0, 0.0; and in P_5 to be 0.0, 51.0, 41.0 and 8.0.

The Eqs. (1′) then read

$$\begin{aligned} 45.4\beta_1+89.5\beta_2 \quad &+100.0\beta_4 \quad =0,\\ 61.6\beta_3 \quad &+51.0\beta_5=0,\\ 54.6\beta_1 \quad &+41.0\beta_5=0,\\ 10.5\beta_2+38.4\beta_3 \quad &+\ 8.0\beta_5=0,\\ &\beta_5=-1. \end{aligned}$$

From the last of these equations, $\beta_5=-1$; from the third, $\beta_1=0.750$; from the second, $\beta_3=0.828$; from the fourth, $\beta_2=-2.27$; and from the first, $\beta_4=1.69$. Then

$$U_Z=0.750\begin{bmatrix}45.4\\0.0\\54.6\\0.0\end{bmatrix}+0.828\begin{bmatrix}0.0\\61.6\\0.0\\38.4\end{bmatrix}+1.69\begin{bmatrix}100.0\\0.0\\0.0\\0.0\end{bmatrix}=\begin{bmatrix}203.0\\51.0\\41.0\\31.8\end{bmatrix}$$

and

$$U_Z=2.27\begin{bmatrix}89.5\\0.0\\0.0\\10.5\end{bmatrix}+\begin{bmatrix}0.0\\51.0\\41.0\\8.0\end{bmatrix}=\begin{bmatrix}203.0\\51.0\\41.0\\31.8\end{bmatrix}.$$

From (4′),

$$u=\sum_{j=1}^{4}u_j$$

and $f_1=0.621$, $f_2=0.156$, $f_3=0.125$ and $f_4=0.097$. If weights other than weight percentages are used as coefficients in Eqs. (1′), the values of the numbers $\beta_1, \cdots, \beta_n$ and $u_1, \cdots, u_n$ may change, but the final weight percentages $f_1, \cdots, f_n$ will remain unaffected.

FORMULAS FOR THE RANGE OF COEXISTENCE OF n PHASES

Suppose that n independent phases $P_1, \cdots, P_n$ have been shown to coexist, it is desired to find, expressed in terms of weight fractions, the composition range of the components $e_1, \cdots, e_n$ which permit this coexistence. Interpreted geometrically, the problem is that of defining, in the $(n-1)$-dimensional composition space, the n-hedron which has the n phases as vertices. Let $f_1, \cdots, f_n$ be the desired weight fractions, then (6) holds. Let $a_{11}, \cdots, a_{nn}$, $\omega_1, \cdots, \omega_n$ have the same meaning as in the previous section. Form the determinant

$$D=\begin{vmatrix}a_{11}\cdots a_{n1}\\ \cdot\\ \cdot\\ \cdot\\ a_{1n}\cdots a_{nn}\end{vmatrix}. \tag{7}$$

The phases form the vertices of a non-degenerate n-hedron, that is, the n phases are independent; thus the n vectors in n-space are independent and $D\neq 0$. Let $\eta=1$ or -1 according as $D>0$ or $D<0$.

The Eq. (6) and the inequalities

$$\begin{vmatrix}\omega_1a_{11}\cdots\omega_1a_{r-1,1} & f_1 & \omega_1a_{r+1,1}\cdots\omega_1a_{n1}\\ \cdot & & \cdot\\ \cdot & & \cdot\\ \cdot & & \cdot\\ \omega_na_{1n}\cdots\omega_na_{r-1,n} & f_n & \omega_na_{r+1,n}\cdots\omega_na_{nn}\end{vmatrix}\cdot\eta\geq 0 \tag{8}$$

for $r=1, \cdots, n$ define the composition region in which $P_1, \cdots, P_n$ coexist.[4]

Since there are n coordinates $f_1, \cdots, f_n$, the Eq. (6) and the inequalities (8) determine a region of an n-dimensional space. If an arbitrary one of the n coordinates f_s is eliminated from the inequalities (8) by means of the Eq. (6) it becomes evident that the region defined by the resulting inequalities is an n-hedron in the $(n-1)$-dimensional composition space expressed in the remaining coordinates $f_1\cdots, f_{s-1}, f_{s+1}, \cdots, f_n$. A simple way to eliminate the variable f_s from the inequalities (8) is, in each determinant, to add to the sth row the sum of all the other rows of that determinant, noting that, because of (6), the element in the rth column and sth row of the final determinant is 1. This yields the relations

$$\begin{vmatrix}\omega_1a_{11} & \cdots\omega_1a_{r-1,1} & f_1 & \omega_1a_{r+1,1} & \cdots\omega_1a_{n1}\\ \cdot & \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot & \cdot\\ \omega_{s-1}a_{1,s-1} & \cdots\omega_{s-1}a_{r-1,s-1} & f_{s-1} & \omega_{s-1}a_{r+1,s-1} & \cdots\omega_{s-1}a_{n,s-1}\\ \sum_{j=1}^{n}\omega_ja_{ij} & \sum_{j=1}^{n}\omega_ja_{r-1,j} & 1 & \sum_{j=1}^{n}\omega_ja_{r+1,j} & \cdots\sum_{j=1}^{n}\omega_ja_{nj}\\ \omega_{s+1}a_{1,s+1} & \omega_{s+1}a_{r-1,s+1} & f_{s+1} & \omega_{s+1}a_{r+1,s+1} & \cdots\omega_{s+1}a_{n,s+1}\\ \cdot & \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot & \cdot\\ \omega_na_{1n} & \cdots\omega_na_{r-1,n} & f_n & \omega_na_{r+1,n} & \cdots\omega_na_{nn}\end{vmatrix}\cdot\eta\geq 0. \tag{9}$$

[4] If weights or weight fractions of the given components are used in describing the phases, that is, if u_{ij} is the weight or weight fraction of the jth component in the ith phase P_i, the numbers u_{ij} may replace the a_{ij} in (7) and ω_ja_{ij} in (8) to obtain equivalent relations describing the region of coexistence. A check on the expansion of the determinants (8) can be made by substituting the weight fractions of the phase P_i for $f_1, \cdots, f_n$ into the inequalities. If $f_{i1}, \cdots, f_{in}$ are the indicated weight fractions, then the left member of the rth inequality is zero if $i\neq r$ and is positive if $i=r$.

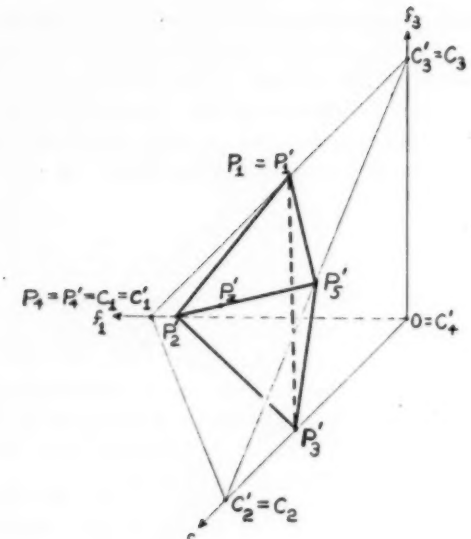


FIG. 2. Under the assumption that P_2 and P_5 coexist at the given temperature and pressure for the composition P_Z, the region of coexistence of P_1, P_2, P_3 and P_5 is shown. Similar regions of coexistence of P_1, P_2, P_4 and P_5 and of P_2, P_3, P_4 and P_5 are to be found, but none will exist for either of the sets P_1, P_2, P_3, P_4 or P_1, P_3, P_4, P_5.

As an illustration of the use of the method, suppose in the first example of the preceding section it has been shown experimentally that the phases P_2 and P_5 coexist. Then P_1, P_2, P_3 and P_5 can coexist. We determine the region of coexistence of these 4 phases.

The determinant D is

$$D=\begin{vmatrix}1&3&0&0\\0&0&2&8\\2&0&0&9\\0&1&3&3\end{vmatrix}=-90<0,$$

so that $\eta=-1$.

The relation (6) becomes

$$f_1+f_2+f_3+f_4=1,$$

and the four inequalities are

$$-\begin{vmatrix}f_1&3\omega_1&0&0\\f_2&0&2\omega_2&8\omega_2\\f_3&0&0&9\omega_3\\f_4&\omega_4&3\omega_4&3\omega_4\end{vmatrix}\geq 0,\quad -\begin{vmatrix}\omega_1&f_1&0&0\\0&f_2&2\omega_2&8\omega_2\\2\omega_3&f_3&0&9\omega_3\\0&f_4&3\omega_4&3\omega_4\end{vmatrix}\geq 0,$$

$$-\begin{vmatrix}\omega_1&3\omega_1&f_1&0\\0&0&f_2&8\omega_2\\2\omega_3&0&f_3&9\omega_3\\0&\omega_4&f_4&3\omega_4\end{vmatrix}\geq 0,\quad -\begin{vmatrix}\omega_1&3\omega_1&0&f_1\\0&0&2\omega_2&f_2\\2\omega_3&0&0&f_3\\0&\omega_4&3\omega_4&f_4\end{vmatrix}\geq 0.$$

When the determinants are expanded these inequalities read

$$-2\omega_2\omega_3\omega_4 f_1-9\omega_1\omega_3\omega_4 f_2+6\omega_1\omega_2\omega_4 f_3+6\omega_1\omega_2\omega_3 f_4\geq 0,$$

$$4\omega_2\omega_3\omega_4 f_1+3\omega_1\omega_3\omega_4 f_2-2\omega_1\omega_2\omega_4 f_3-2\omega_1\omega_2\omega_3 f_4\geq 0,$$

$$-16\omega_2\omega_3\omega_4 f_1-27\omega_1\omega_3\omega_4 f_2+8\omega_1\omega_2\omega_4 f_3+48\omega_1\omega_2\omega_3 f_4\geq 0,$$

$$2\omega_2\omega_3\omega_4 f_1+9\omega_1\omega_3\omega_4 f_2-\omega_1\omega_2\omega_4 f_3-6\omega_1\omega_2\omega_3 f_4\geq 0.$$

To check the expansion, substitute the coordinates of each of the following vectors into the left sides of the above inequalities.

$$W_1=\begin{bmatrix}\omega_1\\0\\2\omega_3\\0\end{bmatrix},\ W_2=\begin{bmatrix}3\omega_1\\0\\0\\\omega_4\end{bmatrix},\ W_3=\begin{bmatrix}0\\2\omega_2\\0\\3\omega_4\end{bmatrix},\ W_5=\begin{bmatrix}0\\8\omega_2\\9\omega_3\\3\omega_4\end{bmatrix}.$$

The coordinates w_{i1}, w_{i2}, w_{i3}, w_{i4} of W_i, substituted[5] for f_1, f_2, f_3, f_4 should make all except the ith one zero and that one should be positive.[6] For example, the coordinates of W_2 make the first, third and fourth left members zero and give the value $10\omega_1\omega_2\omega_3\omega_4$ to the second. This value is positive since the ω_j are atomic weights.

By use of the relation $f_1+f_2+f_3+f_4=1$, an arbitrary one of the variables can be eliminated from the set of inequalities. The resulting inequalities define the interior and boundary of a tetrahedron in the three dimensional space of those coordinates which remain; this tetrahedron is the region of coexistence of the four given phases. See Fig. 2 for a diagram of this tetrahedron.

When the elimination of f_4 is carried out in the example, the inequalities are

$$-2\omega_2\omega_3(\omega_4+3\omega_1)f_1-3\omega_1\omega_3(3\omega_4+2\omega_2)f_2+6\omega_1\omega_2(\omega_4-\omega_3)f_3+6\omega_1\omega_2\omega_3\geq 0,$$

$$2\omega_2\omega_3(2\omega_4+\omega_1)f_1+\omega_1\omega_3(3\omega_4+2\omega_2)f_2-2\omega_1\omega_2(\omega_4-\omega_3)f_3-2\omega_1\omega_2\omega_3\geq 0,$$

$$-16\omega_2\omega_3(\omega_4+3\omega_1)f_1-3\omega_1\omega_3(9\omega_4+16\omega_2)f_2+8\omega_1\omega_2(\omega_4-6\omega_3)f_3+48\omega_1\omega_2\omega_3\geq 0,$$

$$2\omega_2\omega_3(\omega_4+3\omega_1)f_1+3\omega_1\omega_3(3\omega_4+2\omega_2)f_2-\omega_1\omega_2(\omega_4-6\omega_3)f_3-6\omega_1\omega_2\omega_3\geq 0.$$

To check, the W_i are first normalized to weight fractions,[7]

$$f_{11}=\omega_1/(\omega_1+2\omega_3),\ f_{12}=0,\ f_{13}=2\omega_3/(\omega_1+2\omega_3),\ f_{14}=0;$$

$$f_{21}=3\omega_1/(3\omega_1+\omega_4),\ f_{22}=0,\ f_{23}=0,\ f_{24}=\omega_4/(3\omega_1+\omega_4);$$

$$f_{31}=0,\ f_{32}=2\omega_2/(2\omega_2+3\omega_4),\ f_{33}=0,\ f_{34}=3\omega_4/(2\omega_2+3\omega_4);$$

$$f_{51}=0,\ f_{52}=8\omega_2/(8\omega_2+9\omega_3+3\omega_4),\ f_{53}=9\omega_3/(8\omega_2+9\omega_3+3\omega_4),\ f_{54}=3\omega_4/(8\omega_2+9\omega_3+3\omega_4);$$

and substituted into the last inequalities. Only the first three weight fraction coordinates of each point enter into these inequalities. As with the W_i and the

[5] Since there is a positive proportionality constant relating the coordinates of W_i and the weight fraction coordinates $f_{i1}, f_{i2}, f_{i3}, f_{i4}$ of the ith phase, any weight coordinates may be used instead of the weight fraction coordinates in testing these inequalities, due to their homogeneity.

[6] Since the fourth phase was numbered P_5 in the application W_5 will make the left member of the *fourth* inequality positive.

[7] Here the presence of a constant term in the inequalities makes mandatory the use of the weight fractions rather than any set of numbers related to them by a positive proportionality constant.

preceding inequalities, the weight fraction coordinates of the ith one of these points makes the left member of the ith inequality positive and the other left members zero.

APPENDIX I

Derivation of the Formulas for P_Z

Let $A_1, \cdots, A_{n+1}$ be $n+1$ vectors in an n-dimensional space with the property that each n of these vectors form an independent set. Then A_{n+1} is uniquely expressible as a linear combination of the remaining vectors, that is,

$$\alpha_1 A_1 + \cdots + \alpha_n A_n = A_{n+1}, \tag{10}$$

with uniquely determined coefficients $\alpha_1, \cdots, \alpha_n$. Under the hypothesis about the given vectors none of the coefficients is zero, for in that case A_{n+1} would be expressible as a linear combination of less than n of the vectors, giving a dependence among n or less of them. Define $\alpha_{n+1} = -1$ and replace the Eq. (10) by the equation[8]

$$\alpha_1 A_1 + \cdots + \alpha_n A_n + \alpha_{n+1} A_{n+1} = 0. \tag{11}$$

If another linear combination of $A_1, \cdots, A_{n+1}$ equals 0, then the coefficients must be proportional to those in (11). For suppose

$$\gamma_1 A_1 + \cdots + \gamma_n A_n + \gamma_{n+1} A_{n+1} = 0, \tag{12}$$

with coefficients not all zero, then when (11) is multiplied by γ_{n+1} and added to (12), the fact that $\alpha_{n+1} = -1$, makes A_{n+1} be eliminated. If the coefficients in (11) and (12) are not proportional, at least one of the vectors is not eliminated. This implies a dependence among the first n vectors, contrary to hypothesis. Therefore there is exactly one expression of the form (11) satisfying the additional restriction $\alpha_{n+1} = -1$. Moreover none of the α_i are zero, as this would imply a dependence.

If $P_1, \cdots, P_{n+1}$ are $n+1$ phases each n of which are independent and $A_1, \cdots, A_{n+1}$ are vectors representing them, then each n of these vectors is an independent set. Moreover, all coordinates of each vector are positive or zero. It follows from (10) that at least one of the numbers $\alpha_1, \cdots, \alpha_n$ must be positive. Since none of the coefficients in (11) is zero and $\alpha_{n+1} = -1$, the coefficients may be classified according to their signs into two non-void sets. This induces a similar separation of the phases into two sets. If (12) were used to make the separation, the two sets of phases would remain the same, but might be interchanged (an interchange occurs when and only when the proportionality constant is negative). Define[9]

$$A_Z = \sum_{\alpha_i > 0} \alpha_i A_i. \tag{13}$$

Then (11) shows that A_Z is also given by

$$A_Z = \sum_{\alpha_i < 0} -\alpha_i A_i. \tag{14}$$

The vector A_Z is a vector representing the desired composition P_Z. When the A_i are atomic vectors the reduction of A_Z to W_Z, done by multiplying the jth coordinate of A_i by the atomic weight ω_j for each j, is equivalent to introducing a new unit on each of the axes of the n-dimensional space. The subsequent change from weights to weight fractions is equivalent to replacing A_Z by a vector cA_Z with the same direction but extending from the origin to the hyperplane ($(n-1)$-dimensional subspace) $f_1 + \cdots + f_n = 1$. Since two vectors with proportionate coordinates represent the same composition, a point of this hyperplane is thereby chosen as a representative of the composition P_Z.

Moreover, the preceding arguments show that both the composition P_Z and its point representative in the hyperplane are uniquely determined.

The equation $\alpha_{n+1} = -1$, together with the vector equation (11) constitute the Eqs. (1); Eqs. (13) and (14) are the Eqs. (2).

APPENDIX II

Derivation of the Formulas for the Range of Coexistence

In an n-component system let $P_1, \cdots, P_n$ be independent phases which by experiment have been shown to coexist. The problem of determining the range of coexistence resolves itself into that of defining in the hyperplane

$$f_1 + \cdots + f_n = 1 \tag{6}$$

the n-hedron whose vertices are the point representatives of the compositions $P_1, \cdots, P_n$. Since two intersecting, but not coincident, hyperplanes in an n-dimensional space have an $(n-2)$-dimensional intersection, the desired n-hedron may be described as lying in the hyperplane (6) and on the same side as P_r of the hyperplane determined by the origin and the $n-1$ points $P_1, \cdots, P_{r-1}, P_{r+1}, \cdots, P_n$. This must hold for each $r, r = 1, \cdots, n$.

Any linear equation (in n coordinates of the system used to describe the phases) which is satisfied by the coordinates of the origin and each of the compositions except P_r is an equation of the rth one of these hyperplanes. The simplest way to secure a linear equation with this property is to equate to zero the determinant with the coordinates of P_1 in the first column, of P_2 in the second column, etc., and with the coordinates of the variable point in the rth column. Since weight fractions of the various elements are desired, the coordinates used to describe the compositions should be in that system. Since multiplication of the columns of a determinant by positive numbers does not affect the sign or the vanishing of the determinant, it is immaterial whether weights or weight fractions are used in all except the rth column of unknown weight fractions. We are interested in points on the same side of this hyperplane as P_r; that is, points whose coordinates, substituted into the rth column of the determinant, give the determinant the same sign as do the coordinates of P_r.

When the coordinates of P_r are substituted, the determinant has the same sign as the determinant of the atomic coordinates

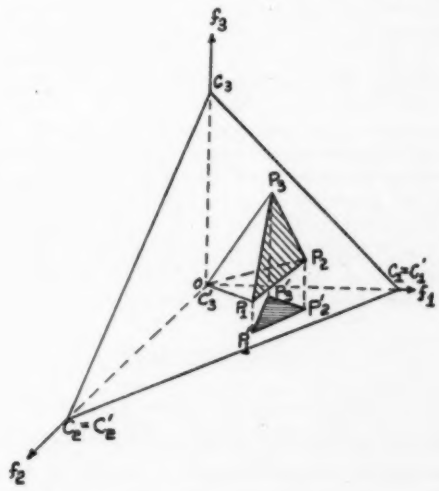


FIG. 3. For a three component system the composition triangle $C_1C_2C_3$, in the three dimensional space, with the points P_1, P_2 and P_3 representing three phases, as used in the present paper, is equivalent to the composition triangle $C_1'C_2'C_3'$ in the f_1f_2-plane with the points P_1', P_2' and P_3' of the customary representation.

[8] Here 0 means the vector with all zero coordinates.

[9] Due to the proportionality between the coefficients in (11) and any equation of the form (12), the use of any such equation instead of (11) in forming (13) and (14) would lead to a vector cA_Z, and, therefore, to a representative of the same composition P_Z.

of the n phases; that is, the same sign as the determinant (7).[10] If this sign is positive, the n determinants described above are positive or zero in the desired n-hedron; if the sign is negative, the determinants are all negative or zero in the n-hedron. This leads to the inequalities (8), which with (6) define the desired n-hedron.

To make the problems capable of simple handling by vectors, a "composition space" described by all n of the components is used. This makes the formulas symmetrical yet does not complicate the geometrical configurations involved. The inequalities (8) describe a "cone" with vertex at the origin in the n-dimensional space. The lateral "surfaces" of the cone are the hyperplanes ($(n-1)$-dimensional spaces in the n-space) whose equations are obtained by replacing the inequality signs by equality signs in the various members of (8). When the Eq. (6) is adjoined to the inequalities (8), the intersection of the cone and the hyperplane (6) is taken. This intersection is an n-hedron in the n-space (if $n=3$, a triangular disk in 3-space). The elimination of the variable f_s from the inequalities (8) by means of (6) is equivalent to projecting the n-hedron onto the fundamental hyperplane determined by $f_1, \cdots, f_{s-1}, f_{s+1}, \cdots, f_n$. Since the hyperplane (6) cuts all n of the axes in the n-dimensional space, the projections of the n-hedron defined by (6) and (8) onto the fundamental hyperplanes are all non-degenerate n-hedra, and it is usually one of the fundamental hyperplanes, with the corresponding projected n-hedron, that is called the composition space. Figure 3 shows a typical case in a three component system, with the projection onto the plane of f_1 and f_2. Here the n-hedron and its projection are triangles. The planes of OP_1P_2, OP_2P_3 and OP_1P_3 are the lateral surfaces of the "cone" (a triangular pyramid in this case). The slant plane cutting the three axes is the plane (6), $f_1+f_2+f_3=1$. In our discussion, this plane is used as the composition plane, rather than the plane of the f_1- and f_2-axes. In view of this fact, all three of the coordinates of each of the points P_1, P_2, P_3 are carried along (in the general case all n coordinates of each of $P_1, \cdots, P_n$). The use of the plane of the f_1- and f_2-axes instead, and the corresponding points P_1', P_2', P_3', is equivalent to projecting our composition plane orthogonally onto the f_1f_2-plane. It is immaterial which plane is used, since the diagrams in the two are totally equivalent.

[10] The determinant of the weight coordinates, being (7) multiplied by the product of the atomic weights of all the n elements involved, has the same sign as the determinant (7). The determinant of weight fraction coordinates equals (7) multiplied by another positive constant. A similar situation holds when the a_{ij} in (7) are molecular coordinates.

On the Diffraction of an Electromagnetic Wave through a Plane Screen

J. W. Miles
Department of Engineering, University of California, Los Angeles, California
(Received December 27, 1948)

Following a discussion of several methods of attack and their application to various particular cases, the problem of the diffraction of an electromagnetic wave through a plane, infinitely thin, perfectly conducting, perforated screen is formulated (in generalized, cylindrical coordinates) in terms of the (generalized Fourier) transform of the tangential electric field in the aperture. The result is an integral equation for this transform, which may also be expressed as an integral equation for the aperture field. The power transferred through the aperture is calculated and cast in a variational form of the Schwinger type, the real and imaginary parts of the reciprocal of the complex power being stationary with respect to first-order variations about the real and imaginary parts of the exact aperture field. An aperture impedance is defined, whose real part (aperture resistance) is the ratio of the aperture scattering cross section to the geometrical cross section, and whose imaginary part is a measure of the standing waves in the neighborhood of the aperture. Moreover, the ratio of the aperture resistance to the impedance of the incident wave is equal to the ratio of the actual power transfer to that predicted by geometrical optics. The real (conductance) and imaginary (susceptance) parts of the aperture admittance are stationary with respect to first-order variations about the real and imaginary parts of the exact aperture field, the former being an absolute minimum in all cases, and the latter only in special cases. A complementary formulation, more suitable to the scattering due to a disk of finite area, is given in terms of the current flowing in the screen. An obstacle admittance, analogous to the aperture impedance, is developed. The two formulations are related by a rigorous form of Babinet's principle (due to Booker, Schwinger, and others). Two more alternative formulations, developed by Copson, are cited. Explicit formulations are given in two-dimensional Cartesian coordinates and in cylindrical polar coordinates. The results are applied to the diffraction of a plane wave through an infinite slit, where the magnetic field is parallel to the slit, and the diffraction of a normally incident plane wave through a circular aperture. The results for the slit compare favorably with the rigorous results computed by Morse and Rubenstein, while the results for the circular hole agree with those of Rayleigh and Bethe in the limit of large wave-length, and with geometrical optics in the limit of small wave-length. The Kirchhoff theory is developed in terms of the aperture conductance and compared with the more exact results. It is found to be very poor in the limit of large wave-length (where the "static" methods developed by Rayleigh are valid) and satisfactory in the limit of small wave-length (where geometrical optics give the transmission factor). The variational formulation provides a convenient continuation between the static and geometric limits and appears to be superior to the Kirchhoff theory for any presumed aperture field.

INTRODUCTION

THE following analysis is directed toward a general formulation of the problem of the diffraction of an electromagnetic wave through an aperture in a plane, perfectly conducting, infinitely thin screen. The complementary problem of the scattering produced by a flat, perfectly conducting disk will also be treated.

The classical method of attack on such problems is due to Kirchhoff and has been widely used in optics.[1] The basic assumption made by Kirchhoff is that the fields in the aperture are those which would exist if the reflecting portion of the screen were not present. While the method is generally adequate for optical problems, where the ratio of the wave-length to the aperture dimensions is quite small, it begins to break down when the wave-length becomes comparable to the aperture

[1] M. Born, *Optik* (Verlag. Julius Springer, Berlin, 1933), p. 151.

dimensions and is completely inadequate when the wave-length is large compared to the aperture dimensions (as in most acoustical problems and many problems at radiofrequencies). A critical discussion of Kirchhoff's method has been given by Baker and Copson.[2]

The problem of producing an exact (closed form) solution to any diffraction problem is difficult, and (with present techniques) often impossible. The best known (successful) example is Sommerfeld's solution of the diffraction due to a half-plane.[3] Macdonald later extended Sommerfeld's solution to the case of an infinite wedge.[4] More recently, Magnus gave an integral equation formulation to Sommerfeld's problem,[5] although he was unable to solve the equation. This integral equation was successfully solved by Copson[6] and, independently, by Schwinger.[7] Schwinger formulated the problem as an inhomogeneous Wiener-Hopf integral equation (Wiener treats only the homogeneous case in his monograph[8]), and pointed out that any diffraction problem involving the incidence of a wave on a number of parallel, conducting cylinders or plates could be so formulated. Following Schwinger's suggestions, Carlson and Heins solved the problem of scattering by a set of staggered plates,[9] and Levine[10] solved the problem of radiation from (or reception by) an open-ended tube.

A diffraction problem closely related to Sommerfeld's is the case of an infinite plane slit. This was first discussed by Schwarzchild,[11] using a method similar to that of Sommerfeld. Strutt solved the problem by separating the wave equation in elliptic coordinates.[12] Numerical results have been given by Morse and Rubenstein.[13] Another problem which may be treated by separation of variables (in oblate spheroidal coordinates) is the diffraction through a circular aperture in a plane screen. The eigenfunctions and eigenvalues have been calculated by Bouwkamp for the acoustical problem.[14] The author was unable to obtain a copy of Bouwkamp's paper until after the present paper had been completed, but it appears that his eigenfunctions may be used in the formulation of the analogous electromagnetic diffraction problem.

Recognizing the difficulties in obtaining closed form solutions, Rayleigh suggested the use of solutions to Laplace's equation for the fields in the aperture and gave approximate solution to several important acoustical and electromagnetic problems.[15,16] More recently, Bethe[17] has utilized the same approximation as Rayleigh to study the diffraction of electromagnetic waves through small circular holes. His solution provided for arbitrary incident waves, whereas Rayleigh treated only the case of plane waves. Bethe's theory has also been applied to holes of other cross section, notably elliptic, by Marcuvitz.[18] Schwinger[7] has also developed very powerful extensions of the use of static solutions in attacking wave guide problems which were essentially two-dimensional.

The present paper proposes to attack the problem by an extension of the methods used (at the M.I.T. Radiation Laboratory) for the treatment of discontinuities in wave guides.[19] The boundary value problem will be formulated as an integral equation, and this in turn, leads to a variational formulation.* The analysis

[2] B. Baker and E. T. Copson, *The Mathematical Foundations of Huygens' Principle* (Clarendon Press, Oxford, 1939).

[3] A. Sommerfeld, *Riemann-Weber's Differential-gleichungen der Physik* (Friedrich Vieweg & Sohn, Braunschweig, 1925), seventh edition, Chapter V. See also Math. Ann. **XLVII**, 317 (1896) and Zeits. f. Math Physik **XLVI**, 2 (1901); W. Pauli has given asymptotic forms of the results, Phys. Rev. **54**, 924–931 (1938). A rather complete bibliography on the problem, together with a discussion of the use of Riemann surfaces in other diffraction problems has been given by H. Bateman, *Electrical and Optical Wave Motion* (Cambridge University Press, London, 1915), Chapter V. Bateman also gives numerous references to various papers on diffraction.

[4] H. M. Macdonald, *Electric Waves* (Cambridge University Press, London, 1902), p. 186.

[5] W. Magnus, Zeits. f. Physik **117**, 168 (1941).

[6] E. T. Copson, Quart. J. Math. **17**, 19 (1945).

[7] J. S. Schwinger, *Seminar on the theory of guided waves*, M.I.T. Radiation Laboratory (1944). (Notes available in Radiation Lab. archives.)

[8] N. Wiener and R. E. A. Paley, *The Fourier Transform in the Complex Domain* (American Mathematical Society. Colloquium Publications **XIX**, New York, 1934), Chapter IV.

[9] J. F. Carlson and A. E. Heins, Quart. App. Math. **IV**, 313–330 (1946); **V**, 82–89 (1947); see also A. E. Heins, Quart. App. Math. **VI**, 157–167 (1948).

[10] H. Levine and J. S. Schwinger, Phys. Rev. **73**, 383–406 (1948).

[11] K. Schwarzchild, Math. Ann. **55**, 177 (1902).

[12] M. J. O. Strutt, *Lamesche, Mathieusche und verwandte Funktionen in Physik und Technik*, Ergeb. d. Math. **1**, 202 (1932); this method was suggested much earlier by W. Wien, Jahresbericht d. deutsch Math. Verein, **15**, 42 (1906); see also B. Sieger, Ann. d. Physik **27**, 626 (1908).

[13] P. M. Morse and P. O. Rubenstein, Phys. Rev. **54**, 895–898 (1938).

[14] C. J. Bouwkamp, *Theoretische en numerieke behandeling van de buiging door een ronde opening* (Dutch), Dissertation, Groningen (1941); published by J. B. Wolten Uitgevers—Maatschapij, N. V. This method was suggested by H. Bateman, *Electrical and Optical Wave Motion* (Cambridge University Press, London, 1915), pp. 90, 97. Recently, and independently, solutions equivalent to Bouwkamp's has been given by R. D. Spence, J. Acous. Soc. Am. **20**, 381–386 (1948) and by A. Storruste and H. Wergeland, Phys. Rev. **73**, 1397 (1948). R. D. Spence and A. Leitner are planning to publish results for the electromagnetic problem, cf., Phys. Rev. **74**, 349 (1948).

[15] Lord Rayleigh, "On the passage of waves through apertures in plane screens and allied problems," Phil. Mag. **XLIII**, 259–272 (1897); *Sci. Papers* **IV**, 283–296 (Cambridge University Press, London, 1903).

[16] Lord Rayleigh, "On the incidence of aerial and electric waves upon small obstacles in the form of ellipsoids or elliptic cylinders, and on the passage of electric waves through a circular aperture in a conducting screen," Phil. Mag. **XLIV**, 28–52 (1897); Sci. Papers **IV**, 305–327 (Cambridge University Press, London 1903).

[17] H. Bethe, Phys. Rev. **66**, 163–182 (1944).

[18] N. Marcuvitz, *Wave guide handbook*, M.I.T. Radiation Lab. Report 41 (January 23, 1945). (To be published by McGraw-Hill Book Company, Inc., New York, as part of the Radiation Lab. series.)

[19] J. W. Miles, Proc. I.R.E. **34**, 728–742 (1946); see also "Contributions to wave-guide theory," Proc. I.R.E. **35**, (April, 1947).

* While the formulation of electromagnetic boundary value problems as integral equations is presumably well known (see, e.g., R. von Mises' treatment in Riemann-Weber, see reference 3, pp. 381–502, particularly *Anwendung der integralgleichungen auf Randwertprobleme*, pp. 439–468; P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (Technology Press, Cambridge, 1947), pp. 198–201, 306–309, 318–323; H. Bateman, "Theory of

closely resembles that recently applied to the problem of acoustic diffraction through a circular aperture.[20] It will also be remarked that a formulation leading to integral equations equivalent to those obtained in the present paper has recently been given by Copson[21] through a rather different approach.

STATEMENT OF PROBLEM

An electromagnetic wave is incident on a screen occupying the plane $z=0$. This screen will be regarded as infinitely thin and consists of an open area, designated as the aperture (σ), and a perfectly conducting area, designated as the obstacle (τ). In general, one of these two areas will be infinite in extent and the other finite, although in some problems (e.g., Sommerfeld's half-plane) both may be infinite. Cylindrical coordinates (u, v, z) will be utilized (e.g., $u=x$, $v=y$ for Cartesian coordinates; $u=r$, $v=\varphi$ for cylindrical polar coordinates). All fields will be assumed to exhibit the harmonic time dependence $\exp(j\omega t)$, and the latter factor will be implicit in all that follows. Then due to the monochromatic nature of the problem, Maxwell's field equations may be written

$$\nabla\times\boldsymbol{H}(u,v,z)=j\beta\zeta\boldsymbol{E}(u,v,z), \tag{1a}$$

$$\nabla\times\boldsymbol{E}(u,v,z)=-j\beta\eta\boldsymbol{H}(u,v,z), \tag{1b}$$

$$\nabla\cdot\boldsymbol{H}(u,v,z)=0, \tag{1c}$$

$$\nabla\cdot\boldsymbol{E}(u,v,z)=0, \tag{1d}$$

where $\boldsymbol{E}$ is the vector electric field, $\boldsymbol{H}$ is the vector magnetomotive field (hereinafter $\boldsymbol{H}$ will be called simply the magnetic field, although this name properly belongs to $\boldsymbol{B}=\mu\boldsymbol{H}$; μ is the permeability), β is the wave number, *viz.*;

$$\beta=\omega/c=2\pi/\lambda. \tag{2}$$

ζ is the characteristic admittance of the (homogeneous) medium, *viz.*,

$$\zeta=\eta^{-1}=(\epsilon/\mu)^{\frac{1}{2}}. \tag{3}$$

ϵ being the dielectric constant and μ the permeability (*MKS* units), c is the characteristic velocity of the medium, *viz.*,

$$c=(\mu\epsilon)^{-\frac{1}{2}}. \tag{4}$$

λ is the wave-length, and ω is the angular frequency. As is well known,[22] a necessary condition for the satisfaction of Maxwell's Eqs. (1) is the satisfaction of the vector wave equation,

$$(\nabla^2+\beta^2)\boldsymbol{E}(u,v,z)=0 \tag{5}$$

plus a similar equation for the magnetic field $\boldsymbol{H}$. In the following analysis, it will be expedient to deal only with the transverse (to the z axis) fields $\mathbf{E}$ and $\mathbf{H}$, which may be separated from $\boldsymbol{E}$ and $\boldsymbol{H}$ by writing

$$\boldsymbol{E}(u,v,z)=\mathbf{E}(u,v,z)+\mathbf{k}E_z(u,v,z), \tag{6a}$$

$$\boldsymbol{H}(u,v,z)=\mathbf{H}(u,v,z)+\mathbf{k}H_z(u,v,z) \tag{6b}$$

where $\mathbf{k}$ is the unit vector along the positive z axis. Moreover, since $\mathbf{k}$ is fixed in direction, E_z and H_z both satisfy the scalar wave equation, and substitution of Eqs. (6) in the vector wave equation (5) yields

$$(\nabla^2+\beta^2)\mathbf{E}(u,v,z)=0 \tag{7}$$

plus a similar equation for $\mathbf{H}(u, v, z)$. It is sufficient to deal with the transverse fields, since the axial fields E_z and H_z may always be determined directly from Eqs. (1).

BOUNDARY CONDITIONS

It is required to find solutions to Eqs. (1) which satisfy the boundary conditions.

$$\lim_{\delta\to 0}\mathbf{E}(u,v,\pm\delta)=0, \quad (u,v)\text{ in }\tau \tag{8a}$$

$$\lim_{\delta\to 0}[\mathbf{E}(u,v,-\delta)-\mathbf{E}(u,v,+\delta)]=0, \quad (u,v)\text{ in }\sigma \tag{8b}$$

$$\lim_{\delta\to 0}[\mathbf{H}(u,v,-\delta)-\mathbf{H}(u,v,+\delta)]=0, \quad (u,v)\text{ in }\sigma \tag{8c}$$

where Eq. (8a) states that the tangential electric field must vanish on the obstacle (τ), while Eqs. (8b, c) state that both the electric and magnetic fields must be continuous across the aperture (σ), although it may sometimes be necessary to keep δ (a small, positive constant) finite during intermediate steps of the analysis. The magnetic field $\mathbf{H}$ will generally be discontinuous across the obstacle due to the current flowing there. Moreover, at the boundary between σ and τ it is not possible to require that the fields be either finite or continuous due to the idealization of the screen as infinitely thin. Bouwkamp[23] has recently criticized the work of Bethe[17] and Copson[21] on the grounds that some boundary condition on the rim of the aperture must be specified in order to effect a unique solution. In discussing Bouw-

integral equations," Proc. London Math. Soc. (2) **4**, 90–115 (1906), and J. App. Phys. **17**, 91–102 (1946)); their power in attacking diffraction problems was brought to the attention of the author (and other workers at the M.I.T. Radiation Lab.) by the elegant (unpublished) work of Schwinger in this field.

[20] A paper "On the diffraction of an acoustic wave through a plane circular aperture," J. Acous. Soc. Am. **21**, 140–141 (1949), was written on this problem by the present writer, but since that time, a more powerful analysis has been given by J. S. Schwinger and H. Levine, Phys. Rev. **74**, 958–974 (1948). While they actually gave results only for acoustic scattering from a circular disk, Levine has informed the writer that they also obtained results for the circular aperture and plan to do further work on the electromagnetic problem. They also cite two previous integral equation approaches to the diffraction problem (L. V. King, Proc. Roy. Soc. (London) **A153**, 1 (1935); A Sommerfeld, Ann. d. Physik **42**, 389 (1942)) not mentioned herein. Levine and Schwinger give variational formulations for the differential scattering cross section (scalar problem).

[21] E. T. Copson, Proc. Roy. Soc. (London) **A186**, 100–118 (1946).

[22] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 392.

[23] C. J. Bouwkamp, A review of Copson's paper (see reference 21), Math. Rev. **8**, 180 (1947).

kamp's criticism, Copson suggests[24] that one would expect that the electric field should be normal to the rim. Unfortunately, it will not generally be possible to satisfy such a condition, since the electric field will become infinite there (except possibly at a re-entrant corner). Insofar as the fields at a distance from the screen are concerned, it does not appear that the boundary condition on the rim is of any great importance, since Bethe's results have received ample experimental verification for sufficiently large ratios of wave-length to aperture diameter.[18] On the other hand, the field in the aperture will probably be rather sensitive to the nature of the boundary condition on the rim, and the assumption of an infinitely thin screen might then be too unrealistic. C. L. Andrews has made a series of remarkable measurements on the electromagnetic fields in a circular aperture,[25] but the calculations for a circular aperture carried out herein or elsewhere (to the author's knowledge) are not sufficiently accurate in the vicinity of the aperture to warrant comparison for wave-lengths small compared to the aperture dimensions.

BASIC SOLUTIONS TO WAVE EQUATION

Following the normal procedure used in many electromagnetic problems (e.g., wave guides), it is convenient to divide the solutions to Maxwell's equations into two sets. These sets will be designated as transverse magnetic (*TM* or superscript 1) and transverse electric (*TE* or superscript 2), corresponding to the absence of z components of the magnetic and electric fields, respectively. The eigenfunctions for the transverse electric field will be taken as $\boldsymbol{\phi}_\lambda(u, v)$, where λ is the eigenvalue, *viz.*,

$$(\nabla^2+\lambda^2)\boldsymbol{\phi}_\lambda(u, v)=0 \tag{9a}$$

$$\boldsymbol{\phi}_\lambda{}^{(2)}=\mathbf{k}\times\boldsymbol{\phi}_\lambda{}^{(1)}. \tag{9b}$$

λ is to be regarded as real and is distributed over a two parametric spectrum.

Vector Transforms

It is expedient to superimpose the elementary solutions by the use of vector transforms[26] over the λ-spectrum. Thus, if $\mathbf{f}(u, v)$ is the electric field in the aperture σ, and $F(\lambda)$ represents its distribution in the λ-spectrum, the transform relations may be written

$$\mathbf{f}(u, v)=\mathbf{T}\{F(\lambda)\}=\int_\lambda dS(\lambda)F(\lambda)\boldsymbol{\phi}_\lambda(u, v) \tag{10a}$$

$$F(\lambda)=\bar{\mathbf{T}}\cdot\{\mathbf{f}(\xi, \eta)\}=\int_\sigma dS(\xi, \eta)\mathbf{f}(\xi, \eta)\cdot\bar{\boldsymbol{\phi}}_\lambda(\xi, \eta), \tag{10b}$$

where T represents the vector transform operator, and the bar denotes the inverse operator (generally the complex conjugate). The integration with respect to the space coordinates (ξ, η) is taken only over σ by virtue of the boundary condition (8a), although, in general, the transform operator implies integration over the entire domain of the function transformed. The integration with respect to λ is carried out over both the *TM* and *TE* spectra.

Total Fields

A solution to the Maxwell's Eqs. (1) for the transverse field on the two sides of the screen $z=0$, due to an incident wave $(\mathbf{E}_i, \mathbf{H}_i)$ in $z<0$, is given by

$$\mathbf{E}(u, v, z)=\begin{Bmatrix}\mathbf{E}_0(u, v, z)\\ 0\end{Bmatrix}+\mathbf{T}\{F(\lambda)\exp[\pm j(\beta^2-\lambda^2)^{\frac{1}{2}}z]\},\ z\lesseqgtr 0 \tag{11}$$

$$\mathbf{H}(u, v, z)=\begin{Bmatrix}\mathbf{H}_0(u, v, z)\\ 0\end{Bmatrix}\mp\zeta\mathbf{k}\times\mathbf{T}\{Y(\lambda)F(\lambda)\exp[\pm j(\beta^2-\lambda^2)^{\frac{1}{2}}z]\},\ z\lesseqgtr 0 \tag{12}$$

where the top and bottom signs refer to negative and positive z, respectively. The $Y(\lambda)$ are designated as the field admittances and are given by

$$Y^{(1)}(\lambda)=[1-(\lambda/\beta)^2]^{-\frac{1}{2}} \tag{13a}$$

$$Y^{(2)}(\lambda)=[1-(\lambda/\beta)^2]^{\frac{1}{2}}. \tag{13b}$$

The branch points $\lambda=\pm\beta$ must be traversed in such a way as to satisfy proper radiation conditions, *viz.*,

$$Y^{(1)}(\lambda)=j[(\lambda/\beta)^2-1]^{-\frac{1}{2}},\quad |\lambda|>\beta \tag{14a}$$

$$Y^{(2)}(\lambda)=-j[(\lambda/\beta)^2-1]^{\frac{1}{2}},\quad |\lambda|>\beta. \tag{14b}$$

$(\mathbf{E}_0, \mathbf{H}_0)$ represents the field corresponding to total reflection of $(\mathbf{E}_i, \mathbf{H}_i)$ in $z<0$. Since $\mathbf{E}_0(u, v, 0)$ vanishes by definition, it is evident that $\mathbf{E}(u, v, z)$, as given by Eq. (11), reduces to the aperture field $\mathbf{f}(u, v)$ at $z=0$.

Transform Integral Equation

Imposing the boundary conditions (8a, c) on Eqs. (11) and (12), respectively, yields

$$\mathbf{T}\{F(\lambda)\}=0,\quad (u, v)\text{ in }\tau \tag{15a}$$

$$\mathbf{k}\times\mathbf{T}\{Y(\lambda)F(\lambda)\}=(2\zeta)^{-1}\mathbf{H}_0(u, v, 0)=\eta\mathbf{H}_i(u, v, 0),\quad (u, v)\text{ in }\sigma \tag{15b}$$

where τ denotes the conducting portion of the screen and σ the aperture. Equation (15) is a "dual" vector integral equation for the determination of $F(\lambda)$. A discussion of this type of (scalar) integral equations is given by Titchmarsh,[27] and it appears that the only form arising from diffraction problems which may be solved with presently known techniques is the Wiener-Hopf type.[6,10]

[24] Personal communication to the author (1948).
[25] C. L. Andrews, Phys. Rev. **71**, 777–781 (1947).
[26] J. W. Miles, Phys. Rev. **74**, 1531 (1948).
[27] E. Titchmarsh, *Theory of Fourier Integrals* (Oxford University Press, London, 1937), pp. 334–349.

In practice, it may be expedient to demand continuity of $\mathbf{H}(u, v, z)$ across a thin sheet bounded by $z=\mp\delta$ in order to retain a dominant factor $\exp(-\delta|\lambda|)$ in the kernel.

Aperture Field Integral Equation

Postmultiplying (vectorially) both sides of Eq. (15b) by $\mathbf{k}$, and taking the inverse transform of both Eqs. (15a, b) yields the results

$$\mathbf{f}(u, v)=0, \quad (u, v) \text{ in } \tau \tag{16a}$$

$$S_\sigma\{\mathbf{g}_\sigma(u, v; \xi, \eta)\cdot\mathbf{f}(\xi, \eta)\}=\mathbf{h}(u, v), \quad (u, v) \text{ in } \sigma \tag{16b}$$

$$\mathbf{g}_\sigma(u, v; \xi, \eta)=\mathbf{T}\{Y(\lambda)\langle\boldsymbol{\phi}_\lambda(\xi, \eta)\rangle\}$$

$$=\int_\lambda dS(\lambda)Y(\lambda)\boldsymbol{\phi}_\lambda(u, v)\langle\boldsymbol{\phi}_\lambda(\xi, \eta)\rangle \tag{17}$$

$$\mathbf{h}(u, v)=\eta\mathbf{H}_i(u, v, 0)\times\mathbf{k}=\mathbf{T}\{H(\lambda)\} \tag{18}$$

where S_σ denotes integration over (ξ, η) in σ. The angular brackets, e.g. $\langle\boldsymbol{\phi}_\lambda\rangle$, denote the complex conjugate. $\mathbf{g}_\sigma$ is the dyadic kernel to the vector integral equation (16b). It may be remarked that the transformation of Eq. (15b) to Eq. (16b) is a generalization of the Faltung (or resultant) theorem for (scalar) Fourier transforms.[28]

It is evident from the preceeding discussion of the boundary conditions at the edge of the aperture that $\mathbf{f}(u, v)$ will generally pass discontinuously from σ to τ, and it is necessary to take Cauchy principal values of the integrals over σ.

Transmitted Power

The complex power transmitted through the aperture is given by

$$P_t=S_\sigma\{\mathbf{k}\cdot\mathbf{E}(u, v, 0)\times\langle\mathbf{H}(u, v, 0)\rangle\}. \tag{19}$$

Transforming the integrand yields

$$P_t=\zeta S_\sigma\{\mathbf{f}(u, v)\cdot\langle\mathbf{h}(u, v)\rangle\}. \tag{20}$$

Now, as an extension of Parseval's theorem for Fourier transforms,[29] it is readily shown that

$$S_{uv}\{\mathbf{f}_1(u, v)\cdot\langle\mathbf{f}_2(u, v)\rangle\}=S_\lambda\{F_1(\lambda)\langle F_2(\lambda)\rangle\} \tag{21}$$

whence Eq. (20) may also be written

$$P_t=\zeta S_\lambda\{F(\lambda)\langle H(\lambda)\rangle\}. \tag{22}$$

On the basis of geometrical optics, Eqs. (19), (20) and (22) reduce to

$$P_t^0=S_\sigma\{\mathbf{k}\cdot\mathbf{E}_i(u, v, 0)\times\langle\mathbf{H}_i(u, v, 0)\rangle\}$$
$$=\zeta Z_i S_\sigma\{|\mathbf{h}(u, v)|^2\}=\zeta Z_i S_\lambda\{|H(\lambda)|\}^2 \tag{23}$$

where $Z_i=Y_i^{-1}$ is defined similarly to $Y(\lambda)$, such that

$$\mathbf{H}_i(u, v, z)=\pm\zeta Y_i[\mathbf{k}\times\mathbf{E}_i(u, v, z)]. \tag{24}$$

[28] See reference 22, p. 313; see reference 27, pp. 1–51.
[29] See reference 27, p. 50.

Aperture Impedance

It is convenient to normalize the transmitted power by introducing the aperture impedance

$$Z_\sigma=S_\sigma\{\mathbf{f}\cdot\mathbf{h}\}/S_\sigma\{|\mathbf{h}|^2\} \tag{25}$$

$$=S_\lambda\{F\bar{H}\}/S_\lambda\{|H|^2\}. \tag{26}$$

The impedance may also be written

$$Z_\sigma=R_\sigma+jX_\sigma=Y_\sigma^{-1} \tag{27a}$$

$$Y_\sigma=G_\sigma+jB_\sigma. \tag{27b}$$

It follows directly from Eqs. (19)–(23) that Z_σ/Z_i is the ratio of the complex power transferred across σ to that predicted by geometrical optics. Since Z_i is real, the ratio of the real powers, i.e., the ratio of the transmission factors, is given by

$$(T/T_0)=(R_\sigma/Z_i). \tag{28}$$

For the case of a plane wave it is seen that R_σ is the ratio of the scattering to the geometric cross section.

Variational Formulation

If $\mathbf{h}$ is substituted in Eq. (20) from Eq. (16b) and the result divided through by the square of Eq. (20) as written, the result is

$$P_t^{-1}=\eta S_\sigma^2\{\langle\mathbf{f}(u, v)\rangle\cdot\mathbf{g}_\sigma(u, v; \xi, \eta)\cdot\mathbf{f}(\xi, \eta)\}/$$
$$|S_\sigma\{\mathbf{f}(u, v)\cdot\langle\mathbf{h}(u, v)\rangle\}|^2. \tag{29}$$

Similarly, from Eqs. (15b) and (22),

$$P_t^{-1}=\eta S_\lambda\{Y(\lambda)|F(\lambda)|^2\}/|S_\lambda\{F(\lambda)\bar{H}(\lambda)\}|^2 \tag{30}$$

and, from Eq. (26),

$$Y_\sigma=S_\lambda\{|H(\lambda)|^2\}S_\lambda\{Y(\lambda)|F(\lambda)|^2\}/$$
$$|S_\lambda\{F(\lambda)\langle H(\lambda)\rangle\}|^2. \tag{31}$$

Equations (29)–(31) are variational in the sense that both their real and imaginary parts are stationary with respect to variations of both the real and imaginary parts of $\mathbf{f}(u, v)$ or $F(\lambda)$ about the exact solutions to the integral equations (15) and (16). The proof is given in the Appendix. Since the real parts of both $Y^{(1)}(\lambda)$ and $Y^{(2)}(\lambda)$ are positive, the real part of P_t^{-1}, i.e., G_σ, is a positive definite form and is therefore an absolute minimum; on the other hand, the imaginary parts of $Y^{(1)}$ and $Y^{(2)}$ are opposite in sign, and B_σ is therefore definite only if the mode structure is entirely TM or TE, and it cannot generally be asserted that B_σ is an absolute minimum.

The foregoing variational equations are extensions of a type of variational equation originated by J. S. Schwinger in unpublished work at the M.I.T. Radiation Laboratory.[7, 19, 30]

[30] Levine and Schwinger have given a more powerful extension of the variational principle, see reference 20, which casts the transmitted (far) field amplitudes in variational form, at least for the scalar problem.

Complementary Formulation

While the foregoing formulation is, in principle, valid for arbitrary σ and τ, it is motivated by the aperture problem (finite σ) and would generally be inappropriate for the obstacle problem (finite τ). In those problems where both σ and τ are infinite, as in the case of the Sommerfeld half-plane or a grating,[31] both approaches might be valuable.

Let $\mathbf{i}(u, v)$ be the vector current density in τ, regarded as a vector current sheet, and let $I(\lambda)$ represent its transform. The boundary condition on the tangential magnetic field is

$$\mathbf{H}(u, v, \mp 0) = \pm \mathbf{k} \times \mathbf{i}(u, v), \quad (u, v) \text{ in } \tau. \tag{32}$$

A solution to Maxwell's equations satisfying the boundary condition (32a) and corresponding to the incident field ($\mathbf{E}_i$, $\mathbf{H}_i$) is given by

$$\mathbf{H}(u, v, z) = \mathbf{H}_i(u, v, z) \pm \mathbf{k} \times \mathbf{T}\{I(\lambda) \exp[\pm j(\beta^2 - \lambda^2)^{\frac{1}{2}} z]\}, \quad z \lesseqgtr 0 \tag{33}$$

$$\mathbf{E}(u, v, z) = \mathbf{E}_i(u, v, z) - \eta \mathbf{T}\{Z(\lambda) I(\lambda) \exp[\pm j(\beta^2 - \lambda^2)^{\frac{1}{2}} z]\}, \quad z \lesseqgtr 0 \tag{34}$$

where the $Z(\lambda)$ are the reciprocals of the $Y(\lambda)$, cf. Eqs. (13), (14).

Satisfaction of the boundary conditions (8a, c) leads to the transform integral equation pair

$$\mathbf{T}\{I(\lambda)\} = 0, \quad (u, v) \text{ in } \sigma \tag{35a}$$

$$\mathbf{T}\{Z(\lambda) I(\lambda)\} = \zeta \mathbf{E}_i(u, v, 0) = \mathbf{e}(u, v), \quad (u, v) \text{ in } \tau \tag{35b}$$

or, alternatively, to the pair

$$\mathbf{i}(u, v) = 0, \quad (u, v) \text{ in } \sigma \tag{36a}$$

$$S_\tau\{\mathbf{g}_\tau(u, v; \xi, \eta) \cdot \mathbf{i}(\xi, \eta)\} = \mathbf{e}(u, v), \quad (u, v) \text{ in } \tau \tag{36b}$$

$$\mathbf{g}_\tau(u, v; \xi, \eta) = S_\lambda\{Z(\lambda) \boldsymbol{\phi}_\lambda(u, v) \langle \boldsymbol{\phi}_\lambda(\xi, \eta) \rangle\}. \tag{37}$$

The two complementary formulations are related by inversion of the respective integral equations, *viz.*,

$$F(\lambda) + \eta Z(\lambda) I(\lambda) = \zeta \bar{\mathbf{T}} \cdot \{\mathbf{e}(u, v)\}. \tag{38}$$

The power reflected from the obstacle is found to be

$$P_\tau = \eta S_\tau\{\mathbf{e}(u, v) \cdot \langle \mathbf{i}(u, v) \rangle\} = \eta S_\lambda\{E(\lambda) \langle I(\lambda) \rangle\} \tag{39}$$

and the obstacle admittance is defined by

$$Y_\tau = S_\tau\{\mathbf{e}(u, v) \cdot \langle \mathbf{i}(u, v) \rangle\} / S_\tau\{|\mathbf{e}(u, v)|^2\} = S_\lambda\{E(\lambda) \langle I(\lambda) \rangle\} / S_\lambda\{|E(\lambda)|^2\} \tag{40}$$

where $E(\lambda)$ is the transform of $\mathbf{e}(u, v)$. Variational equations analogous to Eqs. (29)–(31) may also be written; thus, e.g.,

$$Z_\tau = S_\lambda\{|E(\lambda)|^2\} S_\lambda\{Z(\lambda) |I(\lambda)|^2\} / |S_\lambda\{E(\lambda) \langle I(\lambda) \rangle\}|^2. \tag{41}$$

The real part of Y_τ, i.e., G_τ, is the ratio of the (real) reflected power to that predicted by geometrical optics.

Babinet's Principle

Rigorous formulations of Babinet's principle, originally founded on Kirchoff's theory of diffraction in optics,[32] and found, in that sense, to be incorrect from a more exact viewpoint, have been given by Booker and Copson,[21] Schwinger,[33] and Watson.[34] It can be stated as follows. Consider two screens, complementary in the sense that

$$\sigma^* = \tau \tag{42a}$$

$$\tau^* = \sigma, \tag{42b}$$

and let the fields incident on these two screens satisfy the relations

$$E_i^* = -H_i \tag{43a}$$

$$H_i^* = E_i. \tag{43b}$$

Babinet's principle states that the transmitted fields satisfy the relations

$$E_t + H_t^* = E_i \tag{44a}$$

$$H_t - E_t^* = H_i. \tag{44b}$$

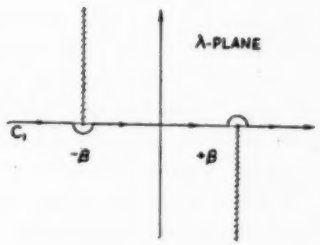


FIG. 1. Path of integration in λ-plane for Eqs. (61) and (62).

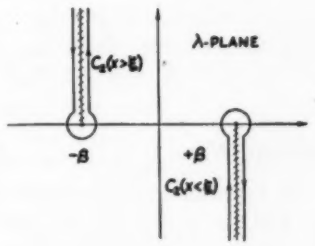


FIG. 2. Deformed contours for Eqs. (61) and (62).

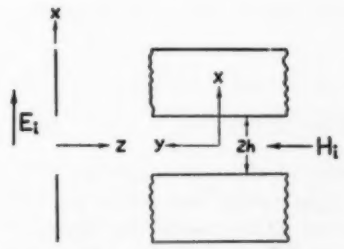


FIG. 3. Cross section of infinite slit with *TM* mode incident.

[31] J. W. Miles, "The diffraction of a plane wave through a grating," Quart. App. Math. 7, 45–64 (1949).

[32] See reference 1, p. 54.

[33] J. S. Schwinger, lectures at M.I.T. Radiation Laboratory (1944), partially in reference 7.

[34] W. Watson, *Physical Principles of Waveguide Transmission* (Oxford University Press, London, 1947).

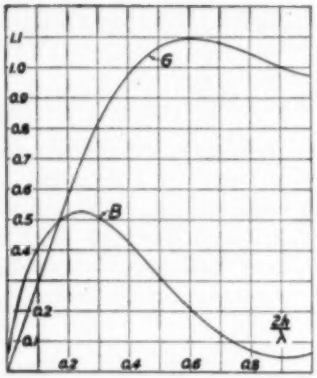


FIG. 4. Conductance and susceptance of slit for normal incidence.

In the present formulation the implied dualities are

$$\mathbf{e}^*(u, v) = \mathbf{h}(u, v) \times \mathbf{k} \tag{45}$$

$$\mathbf{i}^*(u, v) = \mathbf{f}(u, v) \times \mathbf{k} \tag{46}$$

$$\mathbf{g}_\tau^*(u, v; \xi, \eta) = \mathbf{g}_\sigma(u, v; \xi, \eta) \tag{47}$$

$$\eta^* = \zeta \tag{48}$$

$$Z_\tau^* = Y_\sigma. \tag{49}$$

Alternative Formulation

In his papers on diffraction,[15,16] Rayleigh proceeded by applying Green's theorem to the scalar Helmholtz equation. The resulting equations form the starting point of Copson's analysis of the diffraction problem.[21] Their vector analogs have been studied by Schelkunoff,[35] Stratton and Chu,[36] and Smythe.[37] The basic element from which the solutions are constructed in such approaches is the point source, while in the present analysis the basic element is the eigenfunction. Whereas the latter approach is rather less direct, it possesses the distinct advantage of exhibiting the Green's function dependence on (u, v) and (ξ, η) separately, thus facilitating integration (adding, of course, additional integrations over the spectrum coordinates).

Plane Incident Wave

Since any wave may be represented in terms of plane waves, there is no loss in generality in considering the incident wave plane. In the present problem, a convenient representation of a plane wave is given by

$$\mathbf{E}_i(u, v, z) = \boldsymbol{\phi}_i(u, v) \exp(-j\beta z \cos\theta_i) \tag{50}$$

$$\mathbf{H}_i(u, v, z) = \zeta Y_i \mathbf{k} \times \mathbf{E}_i(u, v, z) \tag{51}$$

where θ_i is the polar angle of the wave normal. If the two possible modes of polarization are taken as: *TM* (designated by superscript 1), wherein the magnetic field lies in the plane of incidence, i.e., the plane containing the wave normal and the z axis, and *TE* (designated by the superscript 2, wherein the electric field lies in the plane of incidence, it is found that

$$Y_i^{(1)} = \sec\theta_i \tag{52a}$$

$$Y_i^{(2)} = \cos\theta_i. \tag{52b}$$

It is evident that λ is replaced by $\beta \sin\theta_i$ for a plane wave, cf. Eqs. (13).

The fields corresponding to total reflection of $(\mathbf{E}_i, \mathbf{H}_i)$ are given by

$$\mathbf{E}_0(u, v, z) = -2j\boldsymbol{\phi}_i(u, v) \sin(\beta z \cos\theta_i) \tag{53a}$$

$$\mathbf{H}_0(u, v, z) = 2\zeta Y_i \mathbf{k} \times \boldsymbol{\phi}_i(u, v) \cos(\beta z \cos\theta_i) \tag{53b}$$

whence,

$$\mathbf{h}(u, v) = Y_i \boldsymbol{\phi}_i(u, v) \tag{54}$$

$$\mathbf{e}(u, v) = \zeta \boldsymbol{\phi}_i(u, v). \tag{55}$$

Kirchhoff Results

Inasmuch as the present paper is directed toward an improvement on the Kirchhoff theory, it is only fitting that a comparison of the results predicted by the two theories should be included. Because the Kirchhoff theory is not entirely consistent, there is perhaps some ambiguity as to just how the total power transmission should be calculated. If the power transmission is calculated at the aperture, using the incident fields there, the result is given by Eq. (19) and is labeled as the geometric result. If, on the other hand, the power transmission is calculated by using Green's theorem to compute the radiated fields, but still assuming the incident field in the aperture, the result may be labeled as the Kirchhoff result.

In order to perform the Kirchhoff calculation, $\mathbf{h}$ may be substituted in Eq. (19) from Eq. (16b) to obtain

$$P_t = \zeta RlS_\sigma^2\{\bar{\mathbf{f}} \cdot \mathbf{g}_\sigma \cdot \mathbf{f}\}, \tag{56}$$

since $Rl(\bar{\mathbf{f}} \cdot \mathbf{h}) = Rl(\bar{\mathbf{h}} \cdot \mathbf{f})$. Now, referring to Eqs. (23) and (29) and taking the real part of the latter, Eq. (56) may be written

$$(P_t/P_i^0) = (T/T_0) = (G_\sigma/Z_i) \times [|S_\sigma\{\mathbf{f} \cdot \bar{\mathbf{h}}\}|/S_\sigma\{|\mathbf{h}|^2\}]^2. \tag{57}$$

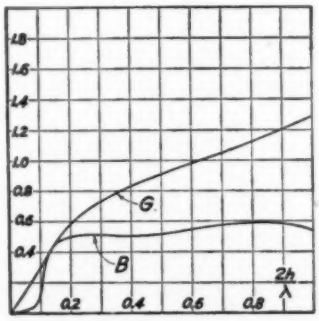


FIG. 5. Conductance and susceptance of slit for oblique incidence (45°).

[35] S. A. Schelkunoff, Phys. Rev. **56**, 308–316 (1939); see also Bell Sys. Tech. J., 92–112 (1936).
[36] J. A. Stratton and L. J. Chu, Phys. Rev. **56**, 99–107 (1939).
[37] W. R. Smythe, Phys. Rev. **72**, 1066–1070 (1947).

The Kirchhoff assumption is

$$\mathbf{f}_K(u, v) = \mathbf{E}_i(u, v, 0), \tag{58}$$

while $\mathbf{h}$ is given by Eq. (18). Hence, Eq. (57) reduces to

$$(T_K/T_0) = (G_\sigma/Y_i), \tag{59}$$

where it is implicit that *G_σ is calculated on the basis of the approximation* (*58*). It should be specifically remarked that Eq. (57) is not a variational formulation for the power, as is Eq. (29), and, in particular, it depends on the magnitude, as well as the distribution, of the field $\mathbf{f}$ inserted. The exact transmission coefficient is given by Eq. (28); thus, for the Kirchhoff approximation (58), the variational formulation gives a result which is the reciprocal of that predicted by the Kirchhoff theory, per se, if the reactance terms (i.e., the standing waves in the vicinity of the aperture) are neglected. The two results will agree only at infinite frequency (where both approach unity), while in other cases the result (28) is superior, by virtue of the variational principle.

Two-Dimensional Problems

Two-dimensional problems are essentially scalar in nature, and their treatment is somewhat special. If the fields are assumed independent of x, the appropriate $\boldsymbol{\phi}_\lambda(x)$ are given by

$$\boldsymbol{\phi}_\lambda^{(1)}(x) = \mathbf{i}(2\pi)^{-\frac{1}{2}} \exp(j\lambda x), \tag{60}$$

and Eqs. (10) reduce to the well-known Fourier transform relations ($\mathbf{i}$ and $\mathbf{j}$ are unit vectors along the x and y axes). The Green's functions are given by Eqs. (17) and (37) as

$$\mathbf{g}_\sigma(x, \xi) = \mathbf{ii}(2\pi)^{-1} \times \int_{-\infty}^{\infty} d\lambda[1-(\lambda/\beta)^2]^{-\frac{1}{2}} \exp[j\lambda(x-\xi)] \tag{61}$$

$$\mathbf{g}_\tau(x, \xi) = \mathbf{jj}(2\pi)^{-1} \times \int_{-\infty}^{\infty} d\lambda[1-(\lambda/\beta)^2]^{\frac{1}{2}} \exp[j\lambda(x-\xi)]. \tag{62}$$

The path of integration, as dictated by the behavior of the solutions at infinity, is shown as C_1 in Fig. 1.

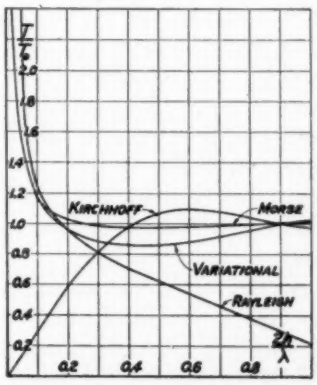


FIG. 6. Slit scattering cross section (relative to geometric) for normal incidence.

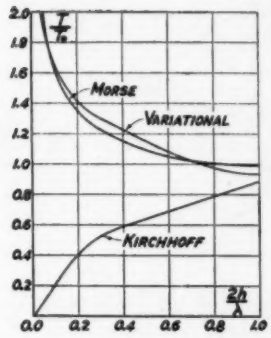


FIG. 7. Slit scattering cross section (relative to geometric) for oblique incidence (45°).

(Allowing z to remain finite would demonstrate absolute convergence.) For $x > \xi$, the integrands are bounded in the upper half-plane, while for $x < \xi$, they are bounded in the lower half-plane. Hence, the contour C_1 may be deformed as shown in Fig. 2, and the integrals may be identified[38] as

$$\mathbf{g}_\sigma(x-\xi) = \mathbf{ii}(\beta/2)H_0^{(2)}(\beta|x-\xi|). \tag{63}$$

$$\mathbf{g}_\tau(x-\xi) = \mathbf{jj}\tfrac{1}{2}|x-\xi|^{-1}H_1^{(2)}(\beta|x-\xi|). \tag{64}$$

The plane wave functions $\boldsymbol{\phi}_i(u, v)$ reduce to

$$\boldsymbol{\phi}_i^{(1)}(x) = \mathbf{i} \exp(-j\beta x \sin\theta_i) \tag{65a}$$

$$\boldsymbol{\phi}_i^{(2)}(x) = \mathbf{j} \exp(-j\beta x \sin\theta_i) \tag{65b}$$

and it is evident that *TM* and *TE* incident waves excites only *TM* and *TE* modes, respectively. (This result is not, in general, true, being characteristic of the essentially scalar nature of the problem.)

Diffraction Through a Slit

Since an exact solution for the problem of diffraction through an infinite slit is available[13] for comparison, an approximate solution was carried out using the present techniques. (It may be remarked that an exact solution to the slit integral equation may be obtained by an expansion in Mathieu functions.) Only the case of a *TM* incident mode was treated, the geometry of the problem being sketched in Fig. 3.

If the approximation

$$\mathbf{f}(x) = \mathbf{i} \exp(-j\beta x \sin\theta_i), \tag{66}$$

together with the Green's function (63), are substituted in Eq. (31), the result may be expressed

$$Y_\sigma(\beta h) = \int_0^{2\beta h} d\eta[1-(\eta/2\beta h)]H_0^{(2)}(\eta) \cos(\eta \sin\theta_i). \tag{67}$$

These results are plotted in Figs. 4 and 5 for $\theta_i = 0$ and 45°, respectively, while the scattering cross sections, cf. Eq. (28), are compared with the exact results of Morse[13] in Figs. 6 and 7. The agreement is evidently

[38] G. N. Watson, *Bessel Functions* (The Macmillan Company, New York 1945), p. 167, Eqs. (6), (7).

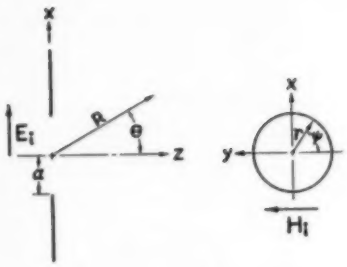


FIG. 8. Circular aperture.

reasonable for both normal and oblique incidence. The Kirchhoff results, cf. Eq. (59), are also plotted for comparison and are evidently completely off for small values of (h/λ).

For small values of βh, the result (67) yields

$$R_\sigma=(\beta h)^{-1}\{1+(2/\pi)^2[1.5-\gamma-\ln(\beta h)]^2\}^{-1}, \quad (68)$$

which is to be compared with the exact (for sufficiently small βh) results given by Rayleigh[15]

$$R_\sigma=(\beta h)^{-1}\{1+(2/\pi)^2[\ln 4-\gamma-\ln(\beta h)]^2\}^{-1}. \quad (69)$$

(γ is Euler's constant.) The result (69) is plotted in Fig. 6 for comparison.

For large values of βh, the asymptotic limits of Eqs. (67) are found to be

$$\lim_{\beta h\to\infty} G_\sigma(\beta h)=\sec\theta_i \quad (70a)$$

$$\lim_{\beta h\to\infty} B_\sigma(\beta h)=0, \quad (70b)$$

whence, cf. Eq. (28)

$$\lim_{\beta h\to\infty} (T/T_0)=1. \quad (71)$$

In this limit both the present and Kirchhoff approaches agree with geometrical optics, as should be expected on physical grounds.

A result, which is considerably superior to the foregoing, may be obtained by substituting the static field,

$$\mathbf{f}(x)=\mathbf{i}[1-(x/h)^2]^{-\frac{1}{2}} \quad (72)$$

in the variational expression (31). The result for normal incidence is

$$Y_\sigma=(2\beta h/\pi)\int_0^{\pi/2} d\theta J_0(\beta h\cos\theta)H_0^{(2)}(\beta h\cos\theta). \quad (73)$$

For small values of βh, the scattering cross section calculated from Eq. (73) is in agreement with Eq. (69), and the general agreement with Morse' results for small βh is better than that given by Eq. (67), although the latter results are superior for larger values of βh.

Polar Coordinates

The appropriate transform eigenfunctions in both Cartesian and polar coordinates have been given elsewhere.[26] The former results will not be used herein, but the latter are required for the following approximate treatment of diffraction through a circular aperture. The eigenfunctions are[26] given by

$$\boldsymbol{\phi}_{\lambda n}^{(1)}(r,\varphi)=(2\pi)^{-\frac{1}{2}}[\mathbf{r}_1 J_n'(\lambda r)+jn\boldsymbol{\varphi}_1(\lambda r)^{-1}J_n(\lambda r)]\cdot\exp[jn\varphi-j(\pi/2)] \quad (74)$$

$$\mathbf{f}(r,\varphi)=\mathbf{T}\{F(\lambda,n)\}=\sum_{n=-\infty}^{\infty}\int_0^\infty \lambda d\lambda F(\lambda,n)\boldsymbol{\phi}_{\lambda n}(r,\varphi) \quad (75a)$$

$$F(\lambda,n)=\bar{\mathbf{T}}\{\mathbf{f}(\xi,\eta)\}=\int_0^\infty \xi d\xi\int_0^{2\pi} d\eta\mathbf{f}(\xi,\eta)\cdot\langle\boldsymbol{\phi}_{\lambda n}(\xi,\eta)\rangle\cdot \quad (75b)$$

The appropriate $\boldsymbol{\phi}_i$ for a plane wave is given by

$$\boldsymbol{\phi}_i^{(1)}(r,\varphi)=[\mathbf{r}_1\cos(\varphi-\varphi_i)-\boldsymbol{\varphi}_1\sin(\varphi-\varphi_i)]\cdot\exp[-j\beta r\sin\theta_i\cos(\varphi-\varphi_i)], \quad (76)$$

where (θ_i,φ_i) are the polar angles of the incident wave normal. The exponential function may be expanded in a Fourier-Bessel series[39] for attacking problems of oblique incidence, but only the case of normal incidence will be treated herein.

Circular Aperture

Perhaps the most interesting diffraction problem from a practical viewpoint is the consideration of a plane wave incident on a circular aperture. (The acoustical problem has been considered elsewhere,[14,20] and Spence and Leitner plan to publish results for the electromagnetic problem.[14]) Approximate expressions for the admittance will be obtained via Eq. (31), using the aperture fields which are valid in the limits of zero and infinite frequency. Using polar coordinates, the aperture is defined by

$$\sigma: 0<r<a. \quad (77)$$

Only the case of normal incidence ($\theta_i=0$) will be considered herein, although approximate results for the case of oblique incidence may be obtained in terms of hypergeometric functions. In this special case $\boldsymbol{\phi}_i^{(1)}$ and $\boldsymbol{\phi}_i^{(2)}$ differ only by a rotation of reference axes. Hence,

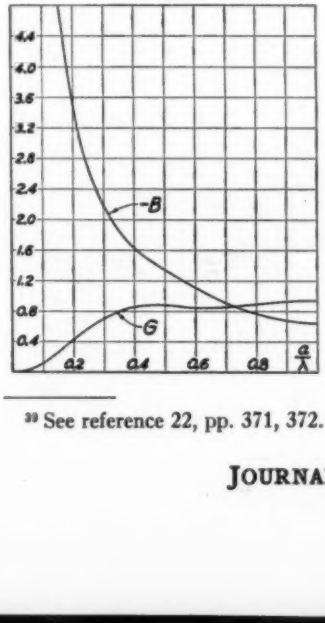


FIG. 9. The conductance and susceptance of a circular aperture for normal incidence of a plane electromagnetic wave.

[39] See reference 22, pp. 371, 372.

it will suffice to take $\varphi_i=0$, whence,

$$\mathbf{h}(r,\varphi)=\mathbf{r}_1\cos\varphi-\hat{\boldsymbol{\varphi}}_1\sin\varphi \tag{78}$$

where the incident electric field is horizontal.

In selecting an approximation for the aperture field, it will be recalled that the static field gives exact results for small βa, whereas the assumption $\mathbf{f}=\boldsymbol{\phi}_i$ is somewhat better for large βa (cf. the comparison between approximate and exact solutions for *TM* diffraction through a slit and for acoustical diffraction through a circular aperture[20]). However, it is important to realize that the assumption $\mathbf{f}=\boldsymbol{\phi}_i$ will actually give entirely incorrect results for small βa in the case where *TE* modes are excited (as they are in diffraction through a circular aperture). To demonstrate this contention, it may first be observed that the *TE* modes dominate the mode structure for small βa by inspecting the Green's functions in the limit. Now the most important component of the *TE* aperture field is H_z (which enters the present solution implicitly, even though it does not appear explicitly in the final equations) since, from Maxwell's equations, it is of the order of the derivatives of the tangential electric field divided by βa, cf. Eq. (1b). Hence, if the aperture field is assumed constant (as it would be for $\mathbf{f}=\boldsymbol{\phi}_i$ and normal incidence), H_z is implicitly neglected, and the nature of the solution is basically modified.** This point was strikingly demonstrated by Bethe's comparison between the exact and Kirchhoff approximations for the scattering cross section for a circular hole.[17]

In calculating the aperture admittance, it will be assumed that the dependence of $\mathbf{f}$ on β is negligible compared to that of $[1-(\lambda/\beta)^2]^{\frac{1}{2}}$ whence, comparing $Y^{(1)}(\lambda)$ and $Y^{(2)}(\lambda)$, cf. Eqs. (13),

$$\beta a Y_a^{(2)}(\beta a)=\int_0^{\beta a} Y_a^{(1)}(z)dz+\lim_{z\to 0}[zY_a^{(2)}(z)]. \tag{79}$$

Hence, the problem is reduced to calculating $Y_a^{(1)}$ for all βa and $Y_a^{(2)}$ only for small βa (it being evident from Eq. (79) that $Y_a^{(2)}\sim(\beta a)^{-1}$).

In the approximate solution of the acoustical problem,[20] it was found that the solution obtained by substituting the static aperture field in the variational results was in excellent agreement with Bouwkamp's exact solution[14] for small values of $\beta a(<1)$, whereas the assumption of a constant aperture field gave rather better results for larger values of βa. However, the latter field gave only about 6–7 percent error for small βa. Now, the acoustical problem corresponds to the *TM* portion of the electromagnetic problem, and no direct analog to the *TE* waves exists. Hence, the diffraction of electromagnetic waves is quite different from the diffraction of sound, at least for small βa.

** The argument might simply have stated that a constant field does not satisfy appropriate boundary conditions, but the matter of primary concern here is that it is more important to satisfy the boundary conditions exactly for small βa, and greater liberty is allowed for larger βa, as is evident from the Kirchhoff theory, which suffices for large βa.

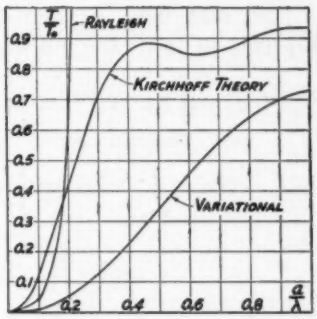


FIG. 10. Scattering cross section (relative to geometric) of a circular aperture for normal incidence of a plane electromagnetic wave.

With the foregoing discussion in mind, the following approximate aperture fields will be used in calculating the transmission through a circular aperture

$$\beta a<<1: \mathbf{E}(r,\varphi,0)=\mathbf{f}(r,\varphi)$$
$$=j(4\beta a/\pi)[1-(r/a)^2]^{\frac{1}{2}}\boldsymbol{\phi}_i(r,\varphi) \tag{80a}$$

$$\beta a\sim 1,\ >>1: \mathbf{f}(r,\varphi)=C\boldsymbol{\phi}_i(r,\varphi). \tag{80b}$$

The first solution was obtained by both Rayleigh[16] and Bethe[17] (replace i by $-j$ in Bethe's results, since he assumes the time factor $\exp(-i\omega t)$) and is exact (for normal incidence only) in the limit of zero βa. It could also be obtained from the present analysis by using the limiting form of the Green's function. For βa not small, Eq. (80b) amounts to the assumption of the incident field in the aperture (and is therefore exact for large βa). It may be remarked that the result (80a) could also be substituted in the variational equation, and the integrations may be carried out in terms of Bessel and Struve functions, as in the acoustical problem,[20] but the results of the latter study scarcely make this worth while, as long as the exact solution will be used for small βa. In calculating the aperture admittance, the two solutions may be easily joined, for the former need enter only in the last term of Eq. (79).

To calculate the aperture impedance for small βa, Eq. (80a), together with Eq. (78), may be substituted directly in Eq. (25). The result is

$$Z_a^{(2)}(\beta a)=j(8/3\pi)\beta a,\quad \beta a<<1. \tag{81}$$

To calculate the remaining terms in $Y_a^{(1)}$, Eq. (31) will be used. Carrying out the integrations yields

$$Y_a^{(1)}=\int_0^\infty \frac{d\lambda}{\lambda}[1-(\lambda/\beta)^2]^{-\frac{1}{2}}J_1^2(\lambda a). \tag{82}$$

Except for a factor of two, Eq. (82) is the radiation admittance (or impedance, depending on definition) of a plane circular piston in an infinite baffle and was calculated by Rayleigh[40] as

$$Y_a^{(1)}(\beta a)=\tfrac{1}{2}[1-(\beta a)^{-1}J_1(2\beta a)]+j(2\beta a)^{-1}\mathbf{H}_1(2\beta a), \tag{83}$$

where $\mathbf{H}_1$ is a Struve function of the first order in the notation of Watson[41] (Rayleigh uses a slightly different notation).

[40] Lord Rayleigh, *Theory of Sound* (Dover Publications, New York, 1945), Vol. II, Section 302.

[41] See reference 38, Section 10.4, p. 328.

$Y_e^{(2)}$ may now be evaluated from Eq. (79), using Eq. (81) to determine the last term. The result is

$$Y_e^{(2)}(\beta a)=\tfrac{1}{2}\Big[1+(\beta a)^{-1}J_1(2\beta a) - (\beta a)^{-1}\int_0^{2\beta a} J_0(z)dz\Big]-(j/2)\Big[(\beta a)^{-1}\mathbf{H}_1(2\beta a) - (\beta a)^{-1}\int_0^{2\beta a}\mathbf{H}_0(z)dz\Big]-j(3\pi/8\beta a). \quad (84)$$

The integral of $J_0(z)$ is tabulated, and may also be evaluated in terms of Struve and Bessel functions.[42] The integral of $\mathbf{H}_0(z)$ is not tabulated, but it entered the problem of the diffraction of sound through a circular aperture when the static aperture field was substituted in the variational expression for the aperture conductance, cf. Eq. (27a) of reference 20, and has been calculated by graphical integration and expansion in a power series.

Adding Eqs. (83) and (84) yields, for the real and imaginary parts of the total aperture admittance

$$G_e(\beta a)=1-(2\beta a)^{-1}\int_0^{2\beta a} J_0(z)dz \quad (85a)$$

$$B_e(\beta a)=(2\beta a)^{-1}\int_0^{2\beta a}\mathbf{H}_0(z)dz-(3\pi/8\beta a). \quad (85b)$$

The power series expansions of the results (85) are readily found to be

$$G_e(\beta a)=\sum_{m=1}^{\infty}\frac{(-)^{m-1}(\beta a)^{2m}}{(2m+1)(m!)^2} \quad (86a)$$

$$B_e(\beta a)=\sum_{m=0}^{\infty}\frac{(-)^m(\beta a)^{2m+1}}{(2m+2)[\Gamma(m+3/2)]^2}. \quad (86b)$$

The transmission coefficient is given by Eq. (28). For small βa, the use of the first few terms in Eqs. (86) yields

$$R_e(\beta a)=(64/27\pi^2)(\beta a)^4\times\{1+[(32/3\pi^2)-(3/20)](\beta a)^2\}=375(a/\lambda)^4\{1+36.8(a/\lambda)^2\}. \quad (87)$$

Multiplying the first term by the area of the aperture, πa^2, gives the "scattering cross section" obtained by Bethe[17] for the circular aperture in the limit of small βa.

The results (85) are plotted in Fig. 9, while the transmission coefficient corresponding to these results is plotted in Fig. 10. It is evident from the figure as well as from Eq. (85), that (T/T_0) approaches unity for large (a/λ), in agreement with Kirchhoff's theory.

Scattering from a Disk

By virtue of Babinet's principle, cf. Eqs. (42)–(49), the results of the last section also represent the impedance (Z_τ) and reflection coefficient (relative to that predicted by geometrical optics) of a flat circular disk of radius a.

[42] See reference 38, p. 752, Table VIII, and footnote.

Formal Solution of Integral Equation

While the primary purpose of the present analysis is the development of approximate expressions for the scattering cross sections, it is important to consider the possibility of solving the integral equations. The Sommerfeld (problem) integral equation may be solved by transforming to parabolic coordinates or by treating it as a Wiener-Hopf type,[5–8] the integral equation for the slit may be solved by transforming to elliptic coordinates and expanding the solution in Mathieu functions, and the integral equation for the circular aperture may be solved by transforming to oblate, spheroidal coordinates and expanding in the appropriate eigenfunctions. Closed form solutions for more general problems appear difficult or impossible at the present time. A formal, implicit solution to the integral equation (16) is obtained by writing

$$\mathbf{f}(u,v)=\sum_n A_n\boldsymbol{\phi}_n(u,v), \quad (88)$$

where $\boldsymbol{\phi}_n$ is an appropriate set of functions. The A_n are given by the solution of the algebraic equations.

$$\sum_n C_{mn}A_n=B_m \quad (89)$$

$$C_{mn}=S_e^2\{\bar{\boldsymbol{\phi}}_m\cdot\mathbf{g}_e\cdot\boldsymbol{\phi}_n\} \quad (90)$$

$$B_m=S_e\{\mathbf{h}\cdot\boldsymbol{\phi}_m\}. \quad (91)$$

Diffracted Fields

Inasmuch as it appears that the methods to be published by Schwinger and Levine[20] will be most appropriate for the calculation of the diffracted fields, this matter will not be taken up herein. It is interesting to remark, however, that the asymptotic fields are determined by the behavior of their transforms in the neighborhood of the origin in the spectrum space. In the present analysis, this fact could be conveniently used to obtain asymptotic representations after (Fourier) transforming with respect to the z coordinate.

APPENDIX

Whereas it was convenient to use dyadic notation in the formulations in the text, it will be convenient in the following to use the standard tensor notation, wherein the presence of a common index denotes summation. The field vectors then become first-order tensors of three components, while the dyadic Green's functions become second-order, symmetric tensors of nine components. Thus, the integral equation (16b) may be written

$$\int_e dS G_{ij}f_j=g_i. \quad (A1)$$

If it is assumed that the function $\mathbf{h}$ is normalized to the condition

$$\int_e \mathbf{g}\cdot\mathbf{g}dS=\int_e g_i\beta_i dS=1, \quad (A2)$$

the aperture impedance is given by Eq. (25) as

$$Z=\int_e f_i\beta_i dS, \quad (A3)$$

and the variational equation for the admittance is

$$DY=\int_\sigma dS\int_\sigma dSf_iG_{ij}f_j \tag{A4}$$

$$D=|Z|^2=\left|\int_\sigma f_ig_i dS\right|^2. \tag{A5}$$

The σ subscripts have been dropped.

It is to be proved that the real and imaginary parts of Y, i.e., G and B, are stationary with respect to first-order variations about the real and imaginary parts of the components f_i. Let the superscripts 1 and 2 denote the real and imaginary parts when appended to G_{ij} and f_i. Separating the real and imaginary portions of (A1), (A2), (A3), and (A)4 yields the equations

$$\int_\sigma dS[G_{ij}{}^1f_j{}^1-G_{ij}{}^2f_j{}^2]=g_i{}^1 \tag{A6a}$$

$$\int_\sigma dS[G_{ij}{}^1f_j{}^2+G_{ij}{}^2f_j{}^1]=g_i{}^2 \tag{A6b}$$

$$R=\int_\sigma dS[f_i{}^1g_i{}^1+f_i{}^2g_i{}^2] \tag{A7a}$$

$$X=\int_\sigma dS[f_i{}^2g_i{}^1-f_i{}^1g_i{}^2] \tag{A7b}$$

$$DG=\int_\sigma dS\int_\sigma dS[f_i{}^1G_{ij}{}^1f_j{}^1+f_i{}^2G_{ij}{}^2f_j{}^2] \tag{A8a}$$

$$DB=\int_\sigma dS\int_\sigma dS[f_i{}^1G_{ij}{}^2f_j{}^1+f_i{}^2G_{ij}{}^2f_j{}^2] \tag{A8b}$$

$$\begin{aligned}D=&\left[\int_\sigma f_i{}^1g_i{}^1dS\right]^2+\left[\int_\sigma f_i{}^1g_i{}^2dS\right]^2+\left[\int_\sigma f_i{}^2g_i{}^1dS\right]^2\\&+\left[\int f_i{}^2g_i{}^2dS\right]^2+2\left[\int_\sigma f_i{}^1g_i{}^1dS\int_\sigma f_i{}^2g_i{}^2dS\right]\\&-2\left[\int_\sigma f_i{}^1g_i{}^2dS\int_\sigma f_i{}^2g_i{}^1dS\right].\end{aligned} \tag{A9}$$

The customary procedure is to set $f_i=f_i{}^{(0)}+\epsilon_i\psi_i$ (where $f_i{}^{(0)}$ represents the exact solution to the integral equation (A6), ψ_i are functions satisfying the same boundary conditions as $f_i{}^{(0)}$ but otherwise arbitrary, and the ϵ_i are arbitrary constants), differentiate G and B with respect to the ϵ_i, set the $\epsilon_i=0$, argue that the ψ_i are arbitrary, replace the resulting integrations with respect to the variables in ψ_i, by the integrands, and seek to prove that

$$\left(\frac{\partial G}{\partial\epsilon_i{}^1}\right)_{\epsilon_i{}^1=0}=\left(\frac{\partial B}{\partial\epsilon_i{}^1}\right)_{\epsilon_i{}^1=0}=\left(\frac{\partial G}{\partial\epsilon_i{}^2}\right)_{\epsilon_i{}^2=0}=\left(\frac{\partial B}{\partial\epsilon_i{}^2}\right)_{\epsilon_i{}^2=0}=0. \tag{A10}$$

An equivalent, although somewhat more formalistic, procedure is to simply vary G and B directly with respect to δf_i. Thus, varying Eqs. (A8) yields

$$D\frac{\delta G}{\delta f_i{}^1}+G\frac{\delta D}{\delta f_i{}^1}=2\int_\sigma dSG_{ij}{}^1f_j{}^1 \tag{A11a}$$

$$D\frac{\delta G}{\delta f_i{}^2}+G\frac{\delta D}{\delta f_i{}^2}=2\int_\sigma dSG_{ij}{}^1f_j{}^2 \tag{A11b}$$

$$D\frac{\delta B}{\delta f_i{}^1}+B\frac{\delta D}{\delta f_i{}^1}=2\int_\sigma dSG_{ij}{}^2f_j{}^1 \tag{A11c}$$

$$D\frac{\delta B}{\delta f_i{}^2}+B\frac{\delta D}{\delta f_i{}^2}=2\int_\sigma dSG_{ij}{}^2f_j{}^2. \tag{A11d}$$

Recalling, cf. Eq. (A5), that $D=R^2+X^2$, it is found that

$$\frac{\delta D}{\delta f_i{}^1}=2Rg_i{}^1-2Xg_i{}^2 \tag{A12a}$$

$$\frac{\delta D}{\delta f_i{}^2}=2Xg_i{}^1+2Rg_i{}^2, \tag{A12b}$$

whence Eqs. (A11) reduce to

$$\tfrac{1}{2}D\frac{\delta G}{\delta f_i{}^1}+G(Rg_i{}^1-Xg_i{}^2)=\int_\sigma dSG_{ij}{}^1f_j{}^1 \tag{A13a}$$

$$\tfrac{1}{2}D\frac{\delta G}{\delta f_i{}^2}+G(Xg_i{}^1+Rg_i{}^2)=\int_\sigma dSG_{ij}{}^1f_j{}^2 \tag{A13b}$$

$$\tfrac{1}{2}D\frac{\delta B}{\delta f_i{}^1}+B(Rg_i{}^1-Xg_i{}^2)=\int_\sigma dSG_{ij}{}^2f_j{}^1 \tag{A13c}$$

$$\tfrac{1}{2}D\frac{\delta B}{\delta f_i{}^2}+B(Xg_i{}^1+Rg_i{}^2)=\int_\sigma dSG_{ij}{}^2f_j{}^2. \tag{A13d}$$

Substituting (A13) in Eqs. (A6), and observing that

$$GX+BR=0 \tag{A14a}$$

$$GR-BX=1 \tag{A14b}$$

yields

$$\frac{\delta G}{\delta f_i{}^1}-\frac{\delta B}{\delta f_i{}^2}=0 \tag{A15a}$$

$$\frac{\delta G}{\delta f_i{}^2}+\frac{\delta B}{\delta f_i{}^1}=0. \tag{A15b}$$

Now δG and δB, due to $\delta f_i{}^1$ or $\delta f_i{}^2$, are not independent, i.e., there exist relations

$$\frac{\delta G}{\delta f_i{}^1}-\alpha\frac{\delta B}{\delta f_i{}^1}=0 \tag{A16a}$$

$$\frac{\delta G}{\delta f_i{}^2}-\beta\frac{\delta B}{\delta f_i{}^2}=0, \tag{A16b}$$

(which may, of course, be determined more explicitly from Eqs. (A13)). Now, since Eqs. (A15) and (A16) represent a complete homogeneous set, it follows that

$$\frac{\delta G}{\delta f_i{}^1}=\frac{\delta G}{\delta f_i{}^2}=\frac{\delta B}{\delta f_i{}^1}=\frac{\delta B}{\delta f_i{}^2}=0, \tag{A17}$$

with the possible exception of the case where the determinant of the set, found to be $1-\alpha\beta$, vanishes; however, a direct investigation of Eqs. (A15) and (A16) for this exceptional case leads also to the results (A17), and the desired proof is completed.

A Method for the Instantaneous Measurement of Velocity and Temperature in High Speed Air Flow

D. G. Marlow, C. R. Nisewanger, and W. M. Cady
U. S. Naval Ordnance Test Station, Physics Division, Pasadena, California

(Received January 3, 1949)

A beam of ultrasonic sound waves is projected across an air stream the velocity and temperature of which are required. From the known ultrasonic frequency and the measured wave-length, the temperature may be determined; and from the slope of the beam boundary, an approximate Mach number may be obtained. The method is applicable to steady-flow problems as in wind tunnels or to transient-flow problems such as the study of shock waves.

INTRODUCTION

THE instrumentation described in this report is the result of a search for suitable methods for determining the instantaneous temperature and velocity in air or other gases. In particular we are concerned with the conditions in the wake of a shock wave produced in a spark shadowgraph. The shocks are generated by the bursting of a diaphragm which closes off

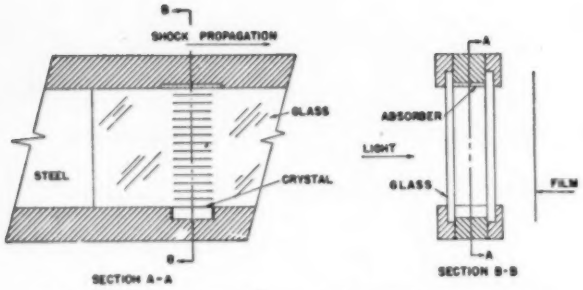


FIG. 1. Schematic arrangement of crystal in shock tube.

the end of a pressure chamber at one end of a long tube of uniform cross section. The shock is transmitted along the tube to a transparent section through which the shadow of the wave may be photographed by collimated light from a synchronized high-intensity spark of one microsecond or less duration. No originality is claimed for this technique. In our version of the device, the tube is rectangular in cross-section, $\frac{3}{4}$ in. wide by 3 in. high, and has a transparent section 10 in. long, the top and bottom of which are made of $\frac{3}{4}$-in. thick steel bars and the sides of $\frac{1}{2}$-in. water-white plate glass. The light from the spark traverses the narrow dimension of the tube, normal to the glass plates.

If a beam of low amplitude sound waves traverses a tube of this nature from, say, bottom to top, one may obtain two useful pieces of information:

1. By measuring the wave-length of the sound waves and combining it with the known frequency, one can obtain the velocity of the waves and hence the temperature, since the velocity depends on temperature only.
2. By measuring the horizontal displacement between successive wave-fronts, one can find the streaming velocity of the gas.

EXPERIMENTAL WORK

Equipment Used

The introduction of a sonic beam into the shock tube was accomplished by placing an X-cut, 344-kc quartz crystal in the bottom of the transparent section. In order to avoid undesired reflections of the shock, it was necessary to inlay the crystal so that its top surface was flush with the bottom of the tube. The general arrangement is shown schematically in the vertical sections, Fig. 1. An absorber of felt, as shown, is inlaid in the top of the shock channel directly over the crystal to absorb the ultrasonic beam and prevent a standing wave pattern. The ultrasonic beam is also indicated.

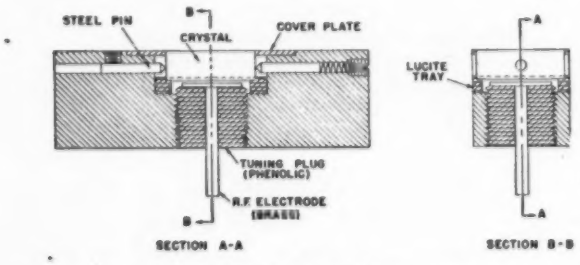


FIG. 2. Details of crystal mounting.

The design of the crystal mounting was complicated by several requirements not encountered in the shadow photography of ultrasonic waves in still air and proved more difficult than we had been led to expect from a study of other workers' techniques.[1]

1. The avoidance of shock reflections at the crystal required the elimination of hold-down fingers, clamps, etc., projecting above the crystal surface.
2. The crystal had to be mounted to withstand the stresses imposed by the shock.
3. The desired crystal outline was square rather than round, in order to produce a beam of uniform optical thickness.
4. The introduction of four air-glass interfaces into the optical path seriously impaired the photographic contrast, probably because of multiple reflections or striae, and necessitated higher driving power and consequent greater danger of crystal fracture.

Requirement 3 was the source of much trouble because of the stress concentration at the corners of the crystal. Several were cracked during the preliminary work until a mounting was devised which damped the corners sufficiently to prevent breakage without seriously reducing the Q. The final design of the mounting is shown in Fig. 2. The crystal is supported by a plate of Lucite which has a channel milled out of the top surface running parallel to the long dimension of the mount. The sides of the channel prevent rotation and sliding crosswise. Through the plate is bored a hole of a diameter such that the crystal rests on its corners, giving them the necessary damping without appreciably spoiling the Q of the crystal as a whole. The longitudinal constraint is supplied by two steel rods pressing against the crystal on opposite sides. One rod is locked to the mount and the other is lightly spring-loaded. The top face of the crystal is silver-plated, the plating being extended down both sides in the form of stripes to make contact with the rods. There is no direct connection to the bottom surface of the crystal, the power being fed capacitatively across the gap between the electrode plate and the quartz. The bakelite plug holding the electrode may be screwed in or out to tune the cavity and thereby improve the Q of the system. In order to prevent severe reflection of the shock as it passes over the crystal, the opening between the latter and the mounting block is closed by an inlaid cover plate of steel fitting closely (0.002 in.) around the crystal perimeter. The entire assembly is inlaid in the bottom of the shock tube so as to present a smooth, continuous surface to the shock wave.

The crystal is driven by a tunable oscillator and power amplifier. If the driving power exceeds about 4 watts the crystal is likely to crack, because the air presents such a poor impedance match that practically all of the power delivered by the amplifier is dissipated in the quartz. The power transfer problem is further aggravated by the mode impurity of X-cut crystals,[2]

[1] Hubbard, Zartman, and Larkin, J. Opt. Soc. Am. **37**, 832 (1947).

[2] L. Bergmann and H. S. Hatfield, *Ultrasonics and Their Scientific and Technical Applications* (John Wiley and Sons, Inc., New York, 1938) pp. 22 and 91.

resulting in local stress concentrations which can easily exceed the elastic limit. Our choice of a square crystal seems to have been a little unfortunate in this respect.

Geometry of the Wave Pattern

The progressive deformation of the supersonic beam during the passage of a shock wave is shown in Fig. 3. The interpretation of these shadowgrams is more easily undertaken with the aid of idealized diagrams such as Fig. 4. A shock travels from left to right with velocity U. In the air ahead of the shock the temperature is T_1, the velocity of sound c_1, and the particle or streaming velocity is zero. Behind the shock these quantities are respectively T_2, c_2, and u_2, all greater than the corresponding quantities ahead of the shock.

Case I. If we consider first those waves which are emitted before the shock arrives at the left end of the crystal, Fig. 4a, we see that as it moves to the right, the elements of the ultrasonic wavefronts are progressively overtaken by air at a higher temperature and hence must move faster than those elements which are still ahead of the shock. Since this acceleration is progressing uniformly across the wavefront, and since the latter must be continuous across the shock, it follows that the left-hand segments of the wavefronts must be tilted up as shown and swept downstream because of the particle velocity u_2. After the shock has passed through the sound beam, the wavefronts will again be unbroken lines but will be tilted and displaced to the right.

Case II. Next we consider those waves which are emitted during the passage of the shock, Fig. 4b. That portion of the wavefront which is behind the shock at the instant of emission will remain parallel to the crystal surface as it travels outward and will exhibit a progressively increasing lead over the right-hand portion. However, since the shock is also moving, the two segments will be joined by a sloping portion because of the mechanism described in the previous paragraph. In addition, the left-hand boundary of the beam will be tilted to the right because of the drift occurring during the interval between the emission of one wave and the next. The locus of the first break in the wavefronts is a straight line leaning to the left. This can be seen by studying the history of the pattern. Wave 1, for example, has just left the crystal and the shock has moved very little in the short interval since the wave was emitted. The horizontal projection of the sloping segment is therefore short. In the case of wave 5, on the other hand, the shock has moved a considerable

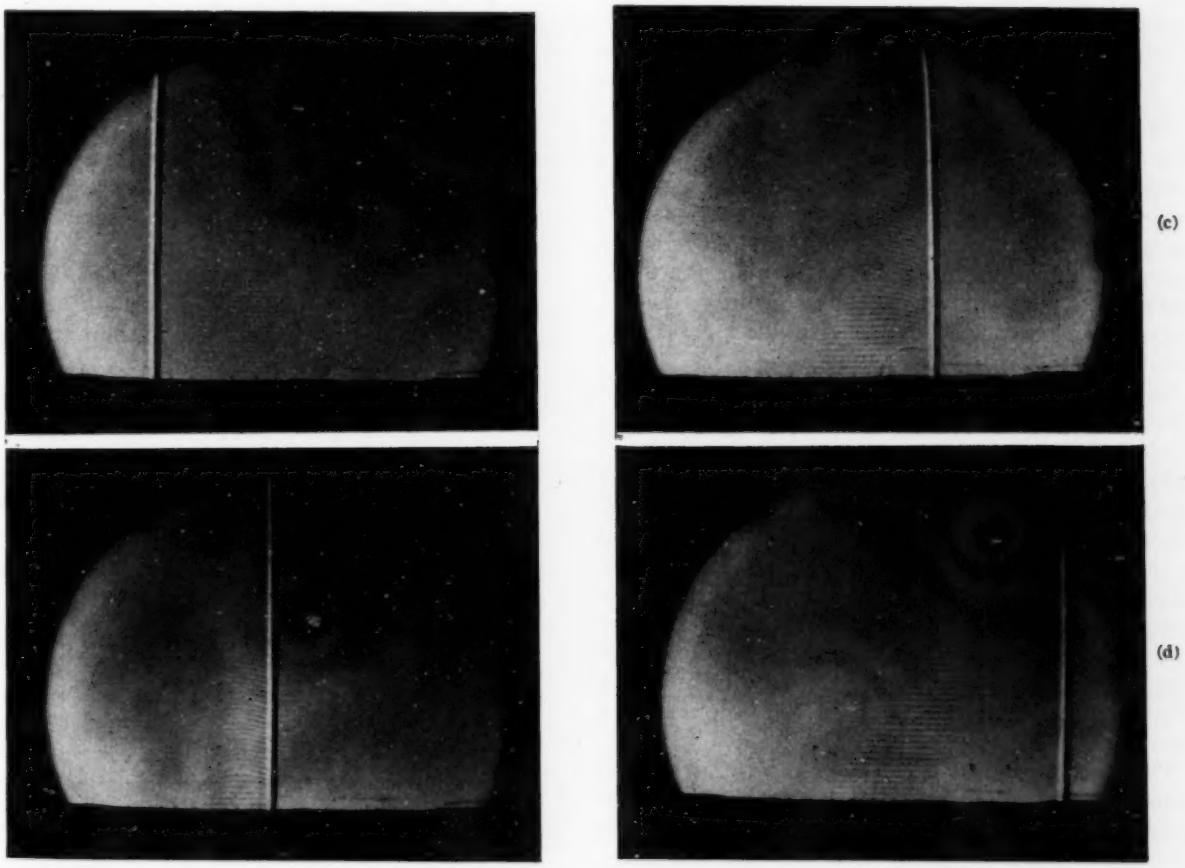


FIG. 3. (a) Ultrasonic field before arrival of shock; (b) ultrasonic field during passage of shock; (c) ultrasonic field approximately 15 μsec. after passage of shock; (d) ultrasonic field approximately 60 μsec. after passage of shock.

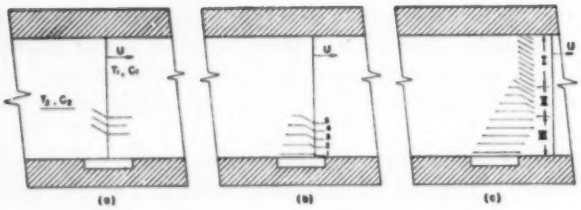


FIG. 4. Idealized diagram of simultaneous phenomena pictured in Fig. 3.

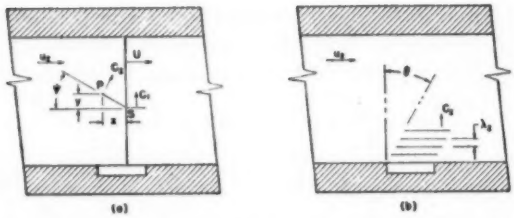


FIG. 5. Diagram defining quantities used in Eq. (1) through (7).

distance to the right of the position it occupied at the instant of emission of the wave, and hence the projection of the associated sloping segment is longer. After the shock has passed the crystal, this group of waves will consist of broken lines having a horizontal segment on the left joined to a sloping segment on the right.

Case III. Finally, waves which leave the crystal after the shock has passed by will remain parallel to the crystal surface as they travel and will be staggered to the right, but will show no other disturbance.

If these phenomena are combined, we get the complete picture of Fig. 4c, which represents the wave pattern after the shock has passed by. The three groups of waves just discussed are bracketed and numbered.

Careful inspection of Fig. 3 will reveal all of these features, somewhat obscured by diffraction, but unmistakable.

The Diffraction Problem

An unsuccessful attempt was made to reduce diffraction by increasing the frequency of the crystal to 1 megacycle. Although some pictures did show a sharper boundary, on the whole the beam was unsatisfactory because of the greater non-uniformity of radiation from the thinner crystal. In order to obtain sharp pictures of the ultrasonic waves at this higher frequency, the spark source was modified to give a flash of 0.1 μsec. duration.

Since, apparently, diffraction could not be sufficiently eliminated by simple means, further experiments were conducted in which diffraction was enhanced so as to form semi-cylindrical waves terminated on a diameter by the bottom of the shock tube. These waves were formed by a thin crystal vibrating in a direction normal to the thickness dimension so that when placed in the shock tube it was essentially a line source of ultrasonic energy. With waves of this form it was hoped to be able to determine the sound velocity by measuring the diameter of the wave shadows and the streaming velocity by measuring the drift of their centers away from the emitting edge.

This technique is similar to that of Carrier and Carlson,[3] who used a spark to generate a pulse disturbance in a wind tunnel which was photographed by Schlieren methods at various intervals after initiation.

[3] G. F. Carrier and F. D. Carlson, Quart. App. Math. 4, 1 (1946).

The crystal had the dimensions: $X=0.5$ mm, $Y=6.4$ mm, $Z=19.1$ mm, and was inlaid in the bottom of the shock tube with the Y direction vertical and the Z direction across the channel. It was driven by an electric field in the X direction, of a frequency such as to excite oscillation along the Y dimension, giving a very narrow emitting surface. This crystal proved to be a rather poor line source because of the impossibility of exciting a uniform edgewise vibration along its entire length, and was abandoned in favor of the original crystal vibrating in a cavity having a narrow slit from which the ultrasonic energy diverged in the clear semi-cylindrical pattern desired.

When a shock wave passed through this pattern, however, it was found that the downstream portions of the ultrasonic waves were so blurred that no information could be obtained from them and it was decided to discontinue work in this direction.

MATHEMATICAL RELATIONSHIPS

If we assume that the propagation of the waves may be adequately described by ray optics, a number of interesting relationships among the various angles shown in Fig. 4 may be derived, only three of which are practically useful.

One of these expresses the angle of tilt of the Case I waves in terms of U, u_2, c_1, and c_2. In Fig. 5a, let y be the vertical advance of the left end of the wavefront P

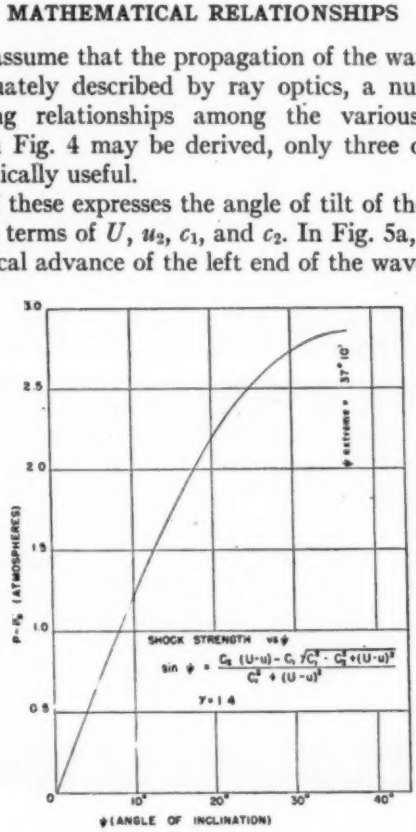


FIG. 6. Plot of shock strength *vs.* ψ.

over the right-hand segment S since the time the shock front first encountered the crystal, and let x be the amount by which P has fallen behind the shock. Then x and y are respectively proportional to the horizontal velocity of P with respect to the shock, and the vertical velocity of P with respect to S. Now the vertical velocity of S is c_1 and of P is

$$v_y = c_2 \cos\psi.$$

Hence

$$y \propto c_2 \cos\psi - c_1. \tag{1}$$

Also, the horizontal velocity of P is

$$v_x = c_2 \sin\psi + u_2 \tag{2}$$

and of the shock is U; hence

$$x \propto U - c_2 \sin\psi - u_2. \tag{3}$$

Therefore

$$y/x = \tan\psi = \frac{c_2 \cos\psi - c_1}{U - u_2 - c_2 \sin\psi}, \tag{4}$$

or

$$U = \frac{c_2 + u_2 \sin\psi - c_1 \cos\psi}{\sin\psi}. \tag{5}$$

If one solves this equation for $\sin\psi$ and makes use of the Rankine-Hugoniot equations, one can make a plot of shock strength *vs.* ψ, as in Fig. 6. It is interesting to note that the curve becomes imaginary for $\psi = 37° 10'$, $P - P_0 = 2.85$. Mathematically, this is an expression of the fact that the discriminant of the solution for $\sin\psi$ vanishes for these conditions. Physically, it means that v_x, Eq. (2), becomes equal to U and that, therefore, all elements of the wavefront which pass through the shock thereafter keep pace with it and do not become visible. For still higher shock pressures, Eqs. (2) to (5) are no longer expressions of a physical situation and the wavefronts simply disappear into the shock, adding their energy to it.

Another quantity of interest is the slope of the left-hand boundary of the Case II and Case III waves, Fig. 4c. Clearly, the horizontal drift velocity of these waves is u_2 and the vertical velocity c_2, hence the angle θ, Fig. 5b, is given by

$$\tan\theta = u_2/c_2 = M \tag{6}$$

where M is the Mach number of the flow behind the shock.

Finally, if the wave-lengths ahead of and behind the shock are measured for those waves which remain parallel to the crystal, one obtains

$$\lambda_2/\lambda_1 = c_2/c_1 = (T_2/T_1)^{\frac{1}{2}}. \tag{7}$$

Of the three quantities ψ, θ, and λ_2/λ_1, the last is the most easily and accurately measured. For low shock pressures, ψ can be measured with fair precision; but θ, which would be a very useful quantity, is at best poorly defined because of diffraction.

Figures 7a and 7b show the progressive changes in a wave in Case I and Case II respectively as the shock passes across the crystal. These pictures should not be confused with Fig. 4, which is a snapshot of a number of different wavefronts at a given instant, while Fig. 7 is a series of snapshots of a particular wavefront taken at different times. The successive positions of the shock are numbered to correspond with the positions of the sonic wavefront.

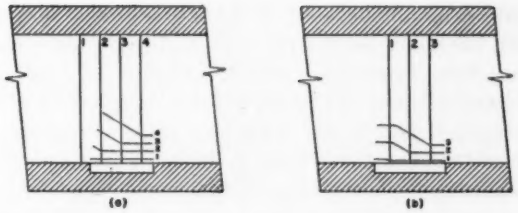


FIG. 7. Progressive deformation of a particular wavefront.

EXPERIMENTAL RESULTS

Figure 3a through 3d is a series of spark shadowgrams showing the passage of a shock through the sonic beam. The wave configurations, blurred because of diffraction at the beam boundaries, are disappointing from the standpoint of Mach number determination, but at least a rough value may be obtained. In order to test the determination of M, the slope of each beam boundary in Fig. 3c was measured four times. The mean value of θ was then used in Eq. (6), giving $M = 0.61$, while the true value, determined from the known shock velocity and the Rankine-Hugoniot equations, was 0.68, an error of 10 percent.

From the standpoint of shock pressure measurements of moderate precision, the picture is a little brighter. Five measurements of ψ on Fig. 3b gave a mean value of 18°.8, corresponding to 2.14 atmospheres shock pressure (Fig. 6). This compares favorably with the value of 2.10 obtained from velocity measurements and the Rankine-Hugoniot equations.

The measurement of temperature can be made with fair precision and reproducibility. As an example, the wave-length was determined from Fig. 3a and again from Fig. 3c by two independent observers measuring with a comparator over a distance of ten wave-lengths. Their independent averages were $\lambda_1 = 1.00$ mm and $\lambda_2 = 1.18$ mm, giving $T_2/T_1 = (\lambda_2/\lambda_1)^2 = 1.41$, whereas the true value determined from the Rankine-Hugoniot equations was 1.44, an error of 3 percent. We feel that this figure represents the precision of the method and is not due to errors in measuring the negatives, since it would require a setting error of $\frac{1}{3}$ wave-length to produce it.

CONCLUSIONS

When this study was begun it was our hope to develop a device which would provide accurate measures of the Mach number; as can be seen from the numerical example, however, it falls short of this goal by a con-

siderable amount because of the impossibility of determining the beam boundaries with sufficient precision.

The measurement of shock pressure through its functional relation to the angle ψ is a by-product of the investigation and is not suggested as a substitute for other methods, such as velocity measurements or piezo gauges. It could have limited applicability, however, in measuring low intensity shocks in regions where their strength is varying and where, consequently, the normally precise velocity method gives no better information.

The real value of the ultrasonic beam method is in its ability to measure temperature accurately and practically instantaneously. For this reason we have called it the ultrasonic thermometer. It may be justly argued that, in order to obtain an accurate figure for λ_2, several wave-lengths must be measured and that, therefore, the temperature calculated from Eq. (7) is only an average over the time of several periods of the wave motion. However, if T_2 should change much in a distance comparable to the length of the crystal face the waves will no longer be parallel to the latter and existence of the temperature gradient can be detected. As long as the waves remain parallel to the emitting surface the temperature gradient is small and the application of Eq. (7) will give good accuracy.

Some Properties of the Ba_2SiO_4 Oxide Cathode Interface*

A. EISENSTEIN**
Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, Massachusetts
(Received January 3, 1949)

Oxide cathodes prepared on a Si-Ni alloy base metal have an interface of barium orthosilicate. The thickness of this layer is measured by means of an x-ray method and found to increase with the life of the cathode and to be of the order of 10^{-3} cm. Measurements of the effective, specific electrical conductivity were made and compared with the conductivity of the coating, (BaSr)O. Both materials exhibit a conductivity-temperature variation characteristic of semiconductors; however, the conductivity of the interface was always less than that of the coating. The interface layer influences the thermionic emission characteristics of the cathode due to an interface voltage developed by the flow of emission current. A retarding-potential method is developed for determining this voltage.

INTRODUCTION

THE oxide cathode interface lying between the base metal and the oxide coating is now regarded as an important cathode parameter although the specific influence of the interface on thermionic emission is not completely clear. Undoubtedly, the type of interface and the conditions under which emission is taken allow the role of the interface to change. The interface region may arise from one of several physical conditions; yet all types are expected to have somewhat similar electronic properties. If the oxide coating is considered to be an impurity semiconductor, the absence of activation centers in the coating immediately adjacent to the base metal may act as a blocking layer[1] to the normal flow of electrons. Poor mechanical bonding[2] of the coating to the base metal gives rise to a thin vacuum layer between the two and may also behave as a blocking layer. Solid-state chemical reactions between the oxide coating and constituents of the base metal frequently give rise to interface compounds whose electrical properties differ from those of the coating. If the conductivity of this interface material is very low, it may act as a blocking layer. These interface forms behave as a series element to the normal flow of electrons from the base metal to the coating; accordingly, a voltage drop[3] should appear across the interface region when a thermionic emission current is taken from the cathode. In this discussion we shall be concerned with the properties of only those interface layers which differ in chemical constitution from the oxide coating and are thus easily detected.

Chemical, spectroscopic, and x-ray diffraction analyses have been used to identify interface compounds, the latter technique giving the most useful information. Rooksby reports[4,5] finding barium aluminate $BaAl_2O_4$ at the interface of cathodes prepared on a 2 percent Al-Ni alloy base metal. Later studies led to the report[6] of barium orthosilicate Ba_2SiO_4 on a 0.4 percent Si-Ni alloy base and barium orthotitanate Ba_2TiO_4 on a 0.23 percent Ti-Ni alloy base. MgO was identified[7] in the interface region of cathodes prepared on a Mg-Ni alloy base. Fineman and Eisenstein[8] studied interfaces formed on Cr-Ni alloys and chromium-plated nickel. Three possible interface compounds were reported, only one of which was stable in air. Although this compound exhibited a close packed hexagonal structure,

* This work has been supported in part by the Signal Corps, the Air Materiel Command, and ONR.
** Now at University of Missouri, Department of Physics.
[1] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, London 1940), p. 176.
[2] J. R. Dillinger, University of Wisconsin, Ph.D. thesis (1947).
[3] A. Eisenstein, Phys. Rev. **71**, 473 (1947).
[4] H. P. Rooksby, J. Roy. Soc. Arts **88**, 318 (1940).
[5] H. P. Rooksby, J. Gen. Elec. Co. **11**, 83 (1940).
[6] H. P. Rooksby, Nature **159**, 609 (1947).
[7] D. A. Wright, Proc. Roy. Soc. Lond. **190**, 394 (1947).
[8] A. Fineman, and A. Eisenstein, J. App. Phys. **17**, 663 (1946).

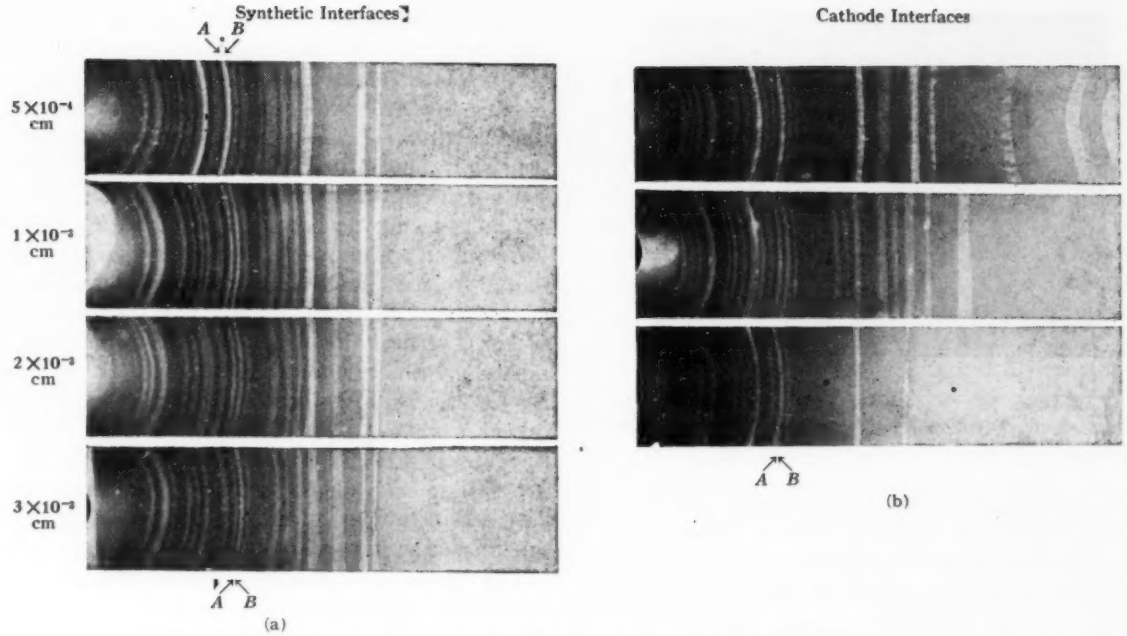


FIG. 1. (a) X-ray diffraction patterns of Ba_2SiO_4 synthetic interface film overlying nickel base metal. Interface thickness 5×10^{-4} cm, 10^{-3} cm, 2×10^{-3} cm, and 3×10^{-3} cm, respectively, reading top to bottom. (b) Diffraction patterns of actual Ba_2SiO_4 cathode interfaces.

none of these interface compounds were chemically identified. X-ray diffraction patterns from the interface region of cathodes prepared on a pure nickel base material show only very weak diffraction lines; hence it seems unlikely that interface compounds are present in this type of cathode.

From this group of interface compounds, barium orthosilicate was selected for a detailed study, since (1) it was chemically stable and easy to synthesize, and (2) it was known to influence the thermionic emission properties of the cathode. A complete study of the influence of the interface layer on the electronic properties of the cathode must include information on the effective electrical conductivity of the interface compound and values of the effective thickness of the interface layer as well as thermionic emission data.

I. INTERFACE FORMATION

The Ba_2SiO_4 interface present in cathodes prepared on a Si-Ni alloy base arises from a chemical reaction between silicon in the base metal and BaO in the coating. Although most oxide coatings contain both barium and strontium, e.g. (BaSr)O, only barium enters into the interface compound.[6,8] When a pure SrO coating is used, a strontium orthosilicate Sr_2SiO_4 interface is formed. The presence of either of these compounds is easily verified by removing the oxide coating[8] and taking x-ray diffraction patterns of the underlying material. O'Daniel and Tscheischwili report[9] the structure of the orthosilicates as being similar to that of K_2SO_4 with the space group D_{2h}^{16}. Ba_2SiO_4 has lattice constants $a=5.7$, $b=10.1$, and $c=7.5$ angstroms. Solid solutions of $(BaSr)_2SiO_4$ are known to occur, although these are not found at the oxide cathode interface.

When the coating is removed from the cathode, the presence of the Ba_2SiO_4 interface is readily apparent by its grey color. The fact that interfaces are easily observed and give good x-ray diffraction patterns,

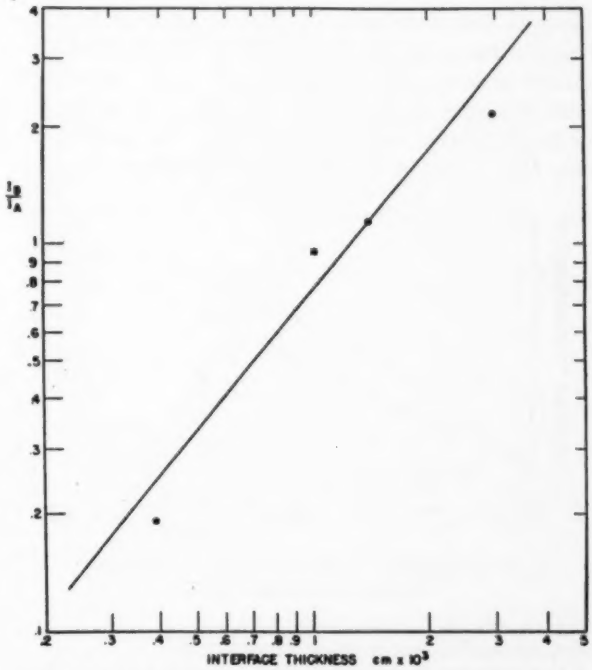


FIG. 2. Ratio of integrated intensities of lines A and B as a function of effective interface thickness.

[9] H. O'Daniel and L. Tscheischwili, Zeits. f. Kristallogr. **104**, 348 (1942).

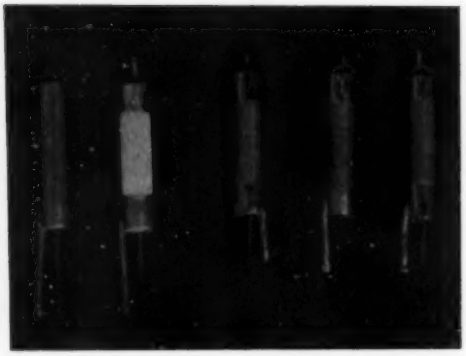


FIG. 3. Reading left to right: uncoated Si-Ni alloy sleeve, normal carbonate coated cathode, barium orthosilicate interface at zero hours, 50 hours, 100 hours.

as well as details of its electrical behavior, led Wright[7] to predict a probable interface thickness range of 10^{-4} to 10^{-5} cm. More recent but unpublished results[10] obtained by Rooksby place this interface thickness at 10^{-3} cm.

An x-ray technique for measuring the thickness of thin crystalline films has been described.[11] This method is based on a comparison of the integrated intensities of x-ray diffraction lines scattered from a thin crystalline surface film and from an underlying crystalline base material. A direct application of this technique to a measurement of the interface thickness is possible only when the absorption coefficient and the crystal structure of the interface material are known. A comparison technique was adopted for evaluating the thickness of the Ba_2SiO_4 interfaces.

Synthetic Ba_2SiO_4 was prepared[12] in a solid-state chemical reaction between $BaCO_3$ and SiO_2 (dehydrated silicic acid) at about 1300°K. This powder in a suitable organic binder was sprayed on the surface of a 3-mm diameter nickel sleeve. The effective thickness of the Ba_2SiO_4 layer was calculated from the weight of the film, the coated area, and the crystalline density:[13]

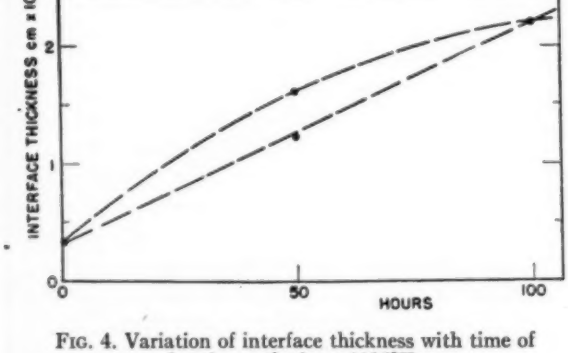


FIG. 4. Variation of interface thickness with time of heating cathode at 1125°K.

[10] D. A. Wright, private communication.
[11] A. Eisenstein, J. App. Phys. 17, 874 (1946).
[12] G. Grube and R. Trucksess, Zeits f. Anorg. u. angew Chemie 203, 84 (1932).
[13] P. Eskola, Am. J. Sci. (5) 4, 331 (1922).

5.7 g/cm³. The true thickness of the film was undoubtedly greater than the calculated effective thickness because of the somewhat lower density of the sprayed coating. No error will occur, however, in the subsequent determination of the actual interface thickness if its density is that of the crystalline solid. X-ray diffraction patterns were taken of the synthetic interface samples with the technique previously described.[8] The sample was rotated continuously during the x-ray exposure to minimize the preferred orientation diffraction spots arising from the large crystal size of the underlying nickel. This technique is essential if accurate measurements of the diffraction line intensity are to be made. Fig. 1a shows a set of x-ray diffraction patterns obtained from synthetic interface samples of effective thickness 5×10^{-4} cm, 1×10^{-3} cm, 2×10^{-3} cm, and 3×10^{-3} cm. Subsequent thickness measurements were based on a comparison of the integrated intensities of the diffraction lines marked *A* and *B*, of the underlying nickel, and of the Ba_2SiO_4 surface film, respectively. These lines occur at nearly the same scattering angle and their respective integrated intensities are easily measured by microphotometering this section of the film. Figure 2 shows a log-log plot of the ratio of the integrated intensities of lines *A* and *B* as a function of the effective interface thickness. An experimental measurement of the ratio of the integrated intensities of these diffraction lines from actual cathode samples, see Fig. 1b, may be referred to Fig. 2 for an immediate determination of the effective interface thickness. A straight line has been drawn through the experimental points as best representing the true behavior since previous work[11] indicated that this log-log function should be linear or have an upward curvature.

If the interface results from a solid-state chemical reaction between constituents of the base metal and the coating, the interface thickness should be a function of the parameters governing the migration of these components. Only one of these, time, has been investigated. Four cathodes were prepared on a 5 percent Si-Ni alloy base metal which had been chemically cleaned and vacuum-fired. Each was coated with $(BaSr)CO_3$,*** to a weight of 10 mg/cm², and each processed in a similar manner with a maximum temperature of 1175°K. One tube was set aside as representing the interface thickness immediately following processing and was called zero hours. Two of the tubes were heated at 1125°K without drawing anode current for 50 hours and the remaining tube for 100 hours. Because of the slight solubility of Ba_2SiO_4 in water, the coatings were removed by dipping in acetone. The underlying interfaces are shown in Fig. 3. For comparison, a normal coated cathode is also shown. A measurement of the interface thickness of each cathode was made using the method described. These thickness values are plotted as a function of the heating time in

*** Raytheon C51-2 coating mix.

Fig. 4. The two tubes at 50 hours indicate the amount of variation to be expected. Unfortunately, this uncertainty does not allow the exact functional relationship to be determined but it seems definite that the interface thickness increases with time and that the order of magnitude of this thickness is 10^{-3} cm. Further studies of this type are being undertaken, for until the factors governing the interface thickness are understood, it is useless to make deductions regarding time effects of the interface on the thermionic emission process.

II. THERMIONIC EMISSION PROPERTIES

The effect of adding reducing impurities to the base metal of an oxide cathode has been examined and discussed by several investigators.[14–16] It is generally believed that the reducing impurities react chemically with the (BaSr)O coating releasing "free" barium which serves as impurity centers in the oxide semiconductor and perhaps also reduces the dipole moment at the vacuum surface of the oxide. This chemical reaction is also consistent with the *F* center model of semiconductor activation which for (BaSr)O is somewhat more satisfying in some respects[17] than is *N*-type activation by excess barium. On the basis of the early experiments,[14–16] one is led to believe that the presence of reducing impurities enhances emission and the presence of oxidizing impurities or vapors reduces the emission as compared to that obtainable from a pure, uncontaminated nickel base metal. It is now becoming evident that many experiments designed to measure the true d.c. emission capabilities of a cathode succeed only in evaluating the state of cathode activity determined in a balance between cathode activation by the reducing impurities and cathode deactivation by vapors from the anode or other tube parts. Jacobs[18] and Hamaker, Gruining, and Aten[19] have clearly indicated the possibility of anode poisoning in such tubes and Dillinger[2] has demonstrated that extremely high, d.c. space-charge-limited emission-current densities may be taken from a pure nickel base cathode, one value being 19 amp./cm² at about 1150°K. The probability of anode poisoning is somewhat reduced in measurements of pulsed emission. Coomes[20] reports that the one-microsecond pulsed emission-current densities which may be obtained from cathodes prepared on a pure nickle base exceeds the emission-current densities from cathodes on the electronic grade *A* nickel[15] containing various reducing impurities. His study presents a comparison of the cathodes sparking current densities which do not necessarily serve as a measure of the cathode's intrinsic activity.

A comparison of the microsecond pulsed voltage-current characteristics of cathodes prepared on a pure nickel and on either a 2 percent or a 5 percent Si-Ni alloy base and tested in the laboratory diode[21] shows two obvious differences (see Fig. 5). For a given temperature the sparking current density, representing the maximum practical limit of emission, is seen to be considerably higher for the pure nickel base cathode than for the alloy base cathode. At 1300°K the sparking current densities were 110 amp./cm² and 10 amp./cm², respectively, for the two types. As the temperature is reduced, the point of departure from the calculated Langmuir-Child space-charge-limited emission line is seen to decrease, following a Richardson type of behavior for the pure nickel base cathode. The alloy base cathode shows no sharp departure from the space-

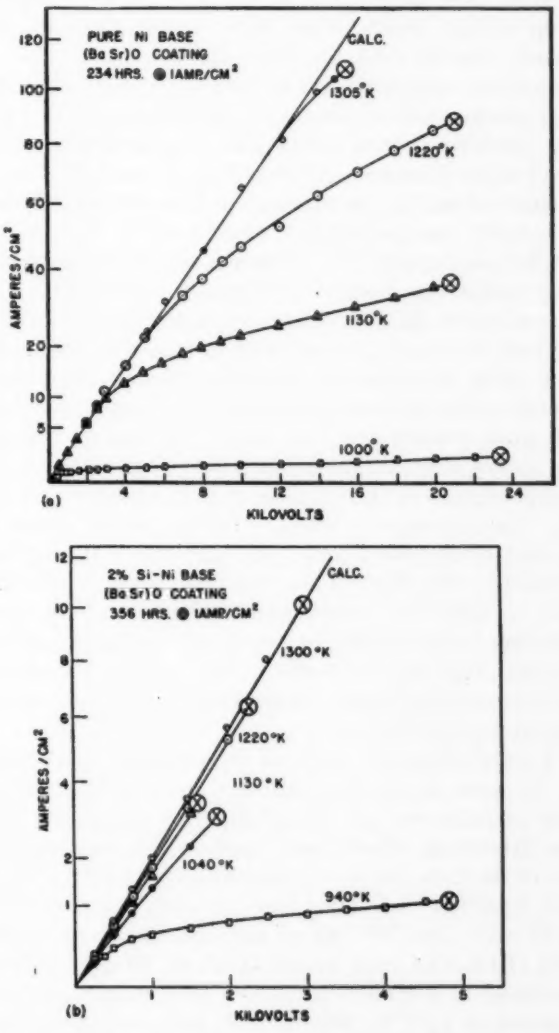


FIG. 5. Voltage-current pulsed-emission characteristics of cathodes prepared on (a) pure nickel base, and (b) 2 percent alloy base. Current plotted on a $\frac{2}{3}$ power scale.

[14] M. Benjamin, Phil. Mag. **20**, 1 (1935).
[15] E. M. Wise, Proc. I.R.E. **25**, 714 (1937).
[16] E. F. Lowry, Phys. Rev. **35**, 1367 (1930).
[17] A. Eisenstein, *Advances in Electronics* (Academic Press Inc., New York, 1948).
[18] Harold Jacobs, J. App. Phys. **17**, 596 (1946).
[19] Hamaker, Gruining, and Aten, Philips Res. Rep. **2**, 171 (1947).
[20] Edward A. Coomes, J. App. Phys. **17**, 647 (1946).
[21] See A. Eisenstein, J. App. Phys. **17**, 654 (1946).

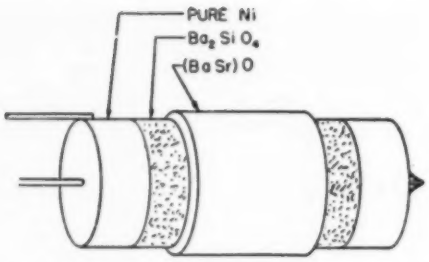


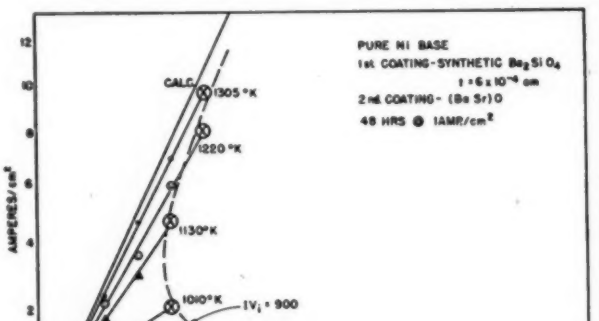
FIG. 6. Sketch showing construction of synthetic interface cathode, (see text).

charge line but rather a change in slope, the apparent point of departure occurring at near zero current density. This "progressive deviation" from the space-charge line at reduced temperatures and the low values of the sparking current density are believed to be characteristic of the interface type cathode and to be directly associated with a sizeable interface voltage drop which occurs when high pulsed currents are passed through the interface layer.

Further confirmation was sought to show that the two characteristics, the progressive deviation and the low sparking current density, are due to the interface layer rather than to a different state of coating activity brought about by the presence of Si in the base metal. A cathode was prepared in which a layer of synthetic Ba_2SiO_4 was applied to a pure nickel base and over this was sprayed a normal coating of $(BaSr)CO_3$. The The effective Ba_2SiO_4 interface thickness was 6×10^{-4} cm and the coating weight was 10 mg/cm². The interface layer extended considerably beyond the oxide-coated width to prevent possible direct contact between the oxide coating and base metal (see Fig. 6). Figures 7a, b show microsecond pulsed voltage-current emission characteristics of this cathode at two stages of cathode life. The cathode was held at 1150°K and d.c. emission current density of 1 amp./cm² was drawn during the life-test period. Both of the emission characteristics are seen to show the progressive deviation and the low sparking current densities previously found in normal Ba_2SiO_4 interface cathodes. Thus it was concluded that these characteristics result from the presence of the silicate interface layer.

A comparison was made of the emission capabilities of the pure nickel base cathode and the Si-Ni alloy base cathode over an extended period of cathode life. Ten laboratory diodes were used in this comparison; five of the cathodes were prepared on pure nickel, three on a 5 percent Si-Ni alloy base, and two on a 2 percent Si-Ni alloy base.**** All of the cathodes were coated with $(BaSr)CO_3$ to a weight of about 10 mg/cm² and processed in a similar manner, with a maximum temperature of 1225°K. Microsecond pulse emission characteristics were recorded at periodic intervals during the life test. Cathodes were operated at a corrected temperature of 1150°K during both the pulse emission measurement and the life-test period. A d.c. emission current density of 1 amp./cm² was drawn from the cathodes during the life test. Figure 8 shows the variation of the sparking current density with the life of the cathode. Wide variations are seen among the tubes, and fluctuations occur in the emission of a single cathode with life. Nevertheless, the sparking current densities of the alloy base cathodes are considerably less than are those of the pure nickel base cathodes. Furthermore, the emission values for the two types of alloy base are nearly the same and show a more rapid decline with increasing cathode life than is found for the pure nickel base. This rapid reduction in the sparking current density may well be due to a change in the interface conductivity or thickness.

No evidence was found for a decay of emission during the pulse although many of the cathodes were tested with both one- and ten-microsecond pulses. This result is contrary to the results reported by Sproull[22] who



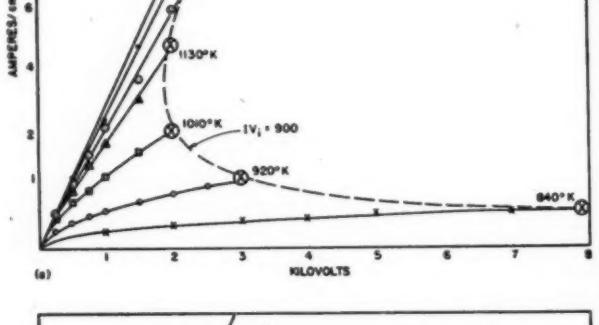


FIG. 7. Voltage-current, pulsed-emission characteristics of synthetic interface cathode (a) at 48 hours, (b) at 310 hours.

**** All of the cathode base materials were kindly supplied by E. M. Wise of the International Nickel Company. The pure nickel was prepared electrolytically and contained no added impurities.

[22] Robert L. Sproull, Phys. Rev. 67, 166 (1945).

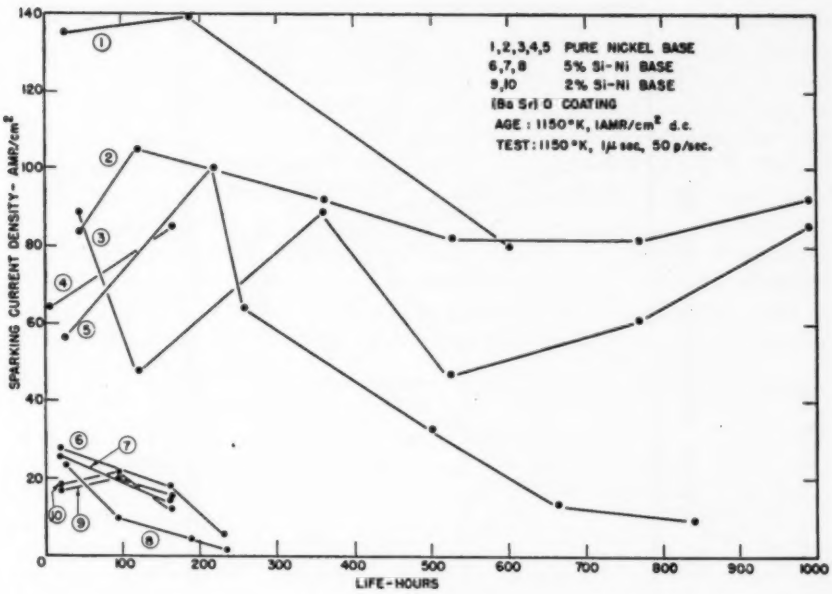

1,2,3,4,5 PURE NICKEL BASE
6,7,8 5% Si-Ni BASE
9,10 2% Si-Ni BASE
(Ba Sr) O COATING
AGE : 1150°K, 1AMR/cm² d.c.
TEST: 1150°K, 1μ sec, 50 p/sec.
SPARKING CURRENT DENSITY - AMR/cm²
LIFE-HOURS

FIG. 8. Variation of sparking-current density with life for pure nickel base cathodes and Si-Ni base cathodes.

found a marked emission decay in the microsecond range for cathodes at 1100°K.

III. CONDUCTIVITY

If the oxide cathode interface gives rise to the progressive deviation type of emission characteristic and controls the sparking current density through the development of a high interface voltage, see Section V, one would expect the electrical conductivity of the interface material to be somewhat less than that of the coating. Numerous experiments[23] have been conducted to evaluate the electrical conductivity of the oxide coating and its temperature variation. Many of these pertain to an unknown state of cathode activity or to bulk samples of the oxides; hence, measurements of the oxide coating conductivity as well as the conductivity of Ba_2SiO_4 seemed desirable.

Measurements of the conductivity of these materials were made in tubes of the type shown in Fig. 9. $(BaSr)CO_3$ or synthetically prepared Ba_2SiO_4 in a suitable organic binder were sprayed on the surface of a MgO ceramic rod.† Half of the desired coating weight was applied and four platinum strips 0.001 in. thick and 2 mm wide were clamped around the sprayed ceramic with the center strips separated by about 15 mm. The remainder of the coating was applied, the platinum bands being left imbedded in the coating. An internal tungsten heater was passed through two axial holes in the ceramic rod. This assembly in cross section is shown in the upper right of Fig. 9. The sample was mounted at the center of a tantalum can which completely surrounded the rod except for two small viewing holes in the side and a slit in the top through which passed the electrical leads. A radiofrequency induction coil surrounded the tube and was used to heat the can which acted as a uniform temperature "furnace" for the sample. The temperature of the can could be made uniform to within 5°C by suitably spacing the turns of the coil. As a check on the temperature, a 0.002-in. tungsten wire was welded to probe lead number 2 in a tube constructed using nickel probe bands. The temperature recorded with this thermocouple was found to be in disagreement by a maximum of 20°C with the assumed blackbody radiation temperature as observed through the side holes with an optical pyrometer.

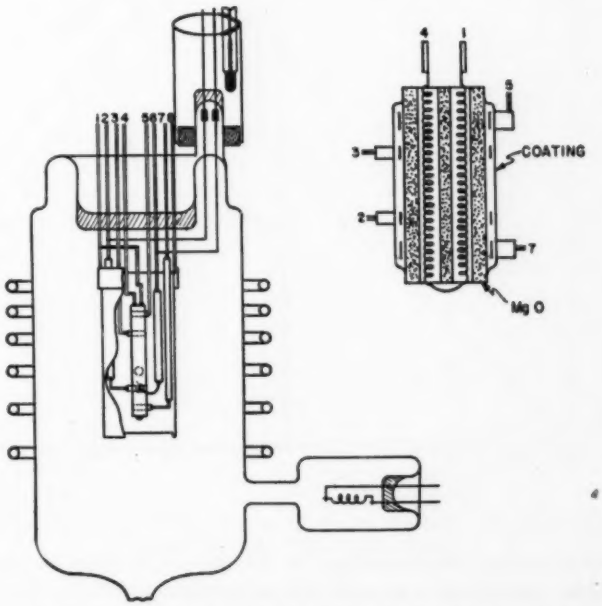
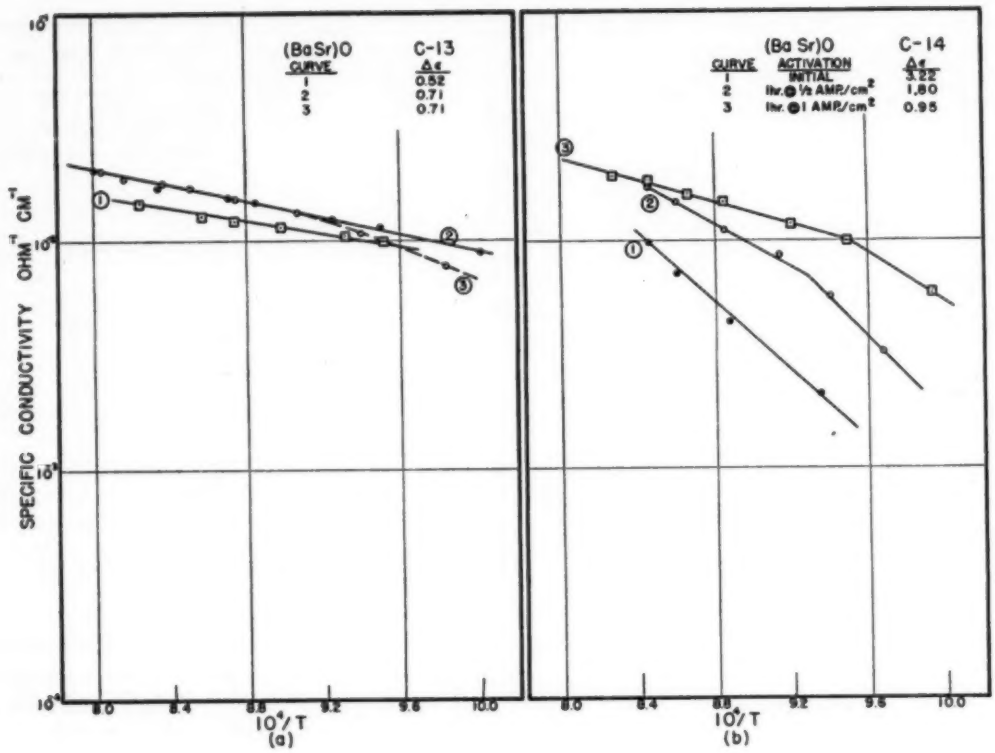


FIG. 9. Tube for measuring electrical conductivity. Enlarged view of the MgO rod is shown at the upper right.

[23] See John P. Blewett, J. App. Phys. **10**, 668 (1939).

† The high purity MgO rods were supplied by A. H. White of the Bell Telephone Laboratories and are a type which were developed for their conductivity studies of oxide coatings. The rod was thoroughly outgassed in vacuum at 1300°K before the coating was applied.



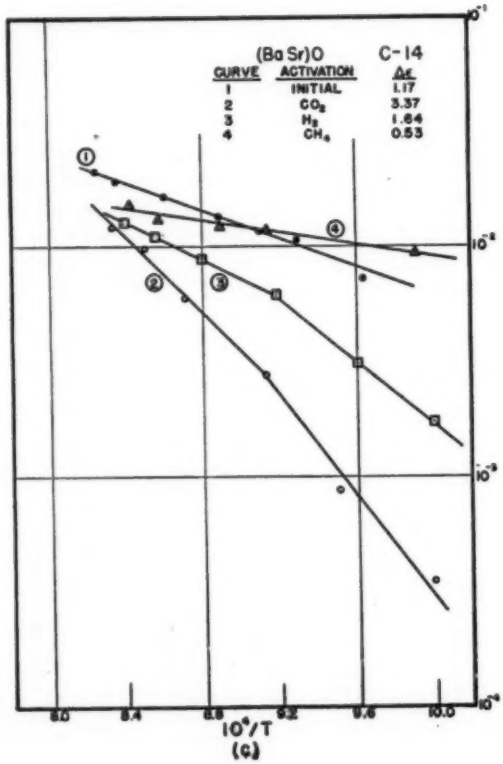


FIG. 10 (a), (b), (c). Temperature variation of the electrical conductivity of (BaSr)O, (see text).

A water bath surrounding the thermocouple press of the tube served as the cold thermocouple junction, (see Fig. 9). For heat treatment in gases, the sample was heated by means of the internal tungsten heater, leads 1 and 4. Conductivity measurements were made by passing a current between probes 5 and 7 and measuring the potential developed between probes 2 and 3 with a potentiometer circuit. The effective specific con-

ductivity of the material between probes 2 and 3 could be evaluated from the equation

$$\sigma = il/Vct, \tag{1}$$

where i is the current passing between probes 5 and 7, l is the distance between probes 2 and 3, V is the potential developed between probes 2 and 3, c is the circumference of the coated ceramic, and t is the coating thickness. Conductivities determined in this manner showed no dependence on the current density except at values approaching 1 amp./cm² which resulted in a heating of the coating. Measurements were generally made with a current density of 10 ma/cm² with the tube continuously exhausted at a pressure of less than 10^{-7} mm. Typical results obtained in the study of equal molar (BaSr)O are shown in Fig. 10. The effective specific conductivity in reciprocal ohm-cm is plotted logarithmically as a function of the reciprocal temperature in degrees Kelvin. Sample *C*-13 was activated at a temperature of 1225°K by passing current densities up to 1 amp./cm², and the conductivity measurementt were taken on three successive days. These curves in sequence comprise Fig. 10a. Over a considerable temperature range the experimental values may be represented by a straight line. This indicates an exponential variation of the type.

$$\sigma = Ke^{-Q/kT}. \tag{2}$$

The similarity between this and the theoretical expression[24] for the temperature dependence of conductivity for semiconductors is easily recognized:

$$\sigma = \frac{4\sqrt{2}}{3} n_b^{\frac{1}{3}} \frac{e^2 l_0}{h^{\frac{1}{3}}} (2\pi m^* kT)^{\frac{1}{2}} e^{-\Delta\epsilon/2kT}, \tag{3}$$

where n_b = density of impurity centers, l_0 = mean free path, $\Delta\epsilon$ = energy gap separating the bottom of the conduction band and the impurity levels, and e, h, m^*, k, and T have their usual significance. The energy Q, obtainable from the slopes of the curves of Fig. 10 is related to the thermal activation energy $\Delta\epsilon$ by

$$Q = \tfrac{1}{2}\Delta\epsilon. \tag{4}$$

Values of $\Delta\epsilon$ of 0.52, 0.71, and 0.71 electron volts were thus obtained, respectively, from the three curves of Fig. 10a.††

Activation of sample *C*-14 by the passage of a high current density is seen in Fig. 10b. An upward shift of the curves accompanied by a decrease of slope attends this activation. In the initial, unactivated state the thermal activation energy is 3.22 ev and decreases to 0.95 ev after the passage of a 0.5 amp./cm² current density for one hour and 1 amp./cm² for one hour. Some time later this sample showed an activation energy of 1.17 ev curve 1, Fig. 10c. The sample was exposed to a few millimeters pressure††† of CO_2 and heated by means of the internal tungsten coil to 1100°K for a period of 5 minutes. Following this treatment, the tube was quickly exhausted and curve 2 was taken. Similar treatements for like time and pressure conditions in H_2 and CH_4 resulted in curves 3 and 4, respectively. Activation and deactivation in the emission from oxide cathodes in CH_4 and CO_2 were discussed by Prescott and Morrison[25] in terms of a chemical reduction of BaO or an oxidation of free barium in the coating.

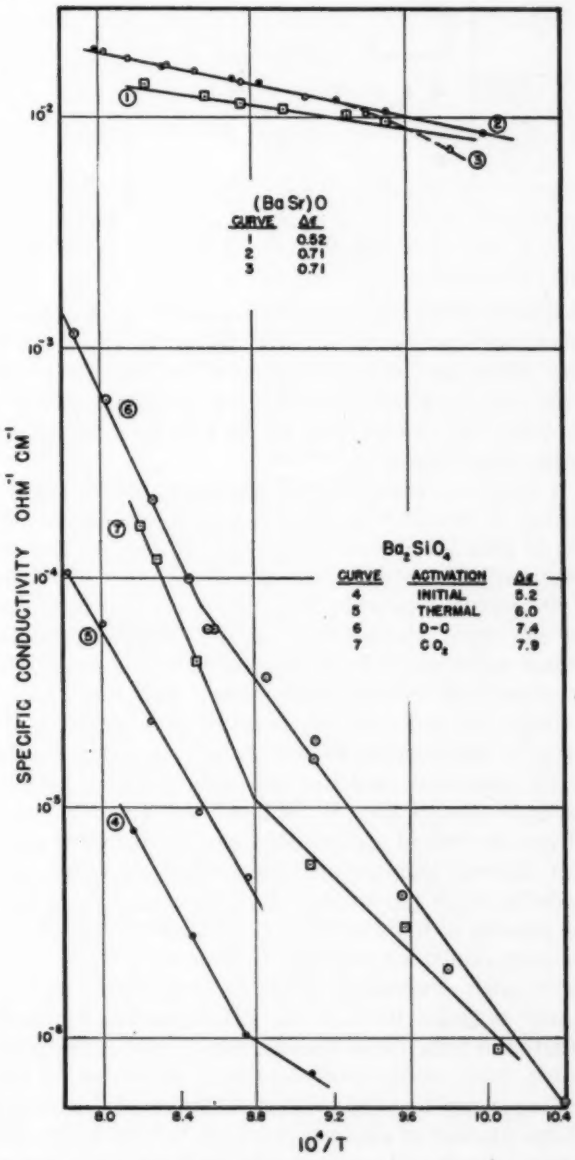


FIG. 11. Comparison of the tempersture variation of the electrical conductivities of (BaSr)O and Ba_2SiO_4.

[24] Frederick Seitz, *Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), p. 191.

†† Conyers Herring has recently shown that if a semiconductor model is chosen in which the number of impurity centers exceeds the number of available electrons, the factor of ½ in Eq. (4) should, for a (BaSr)O coating, be replaced by a value near unity.

††† CO_2, H_2, and CH_4 gases used in this experiment were Airco products procured in glass flasks.

[25] C. H. Prescott and J. Morrison, J. Am. Chem. Soc. **60**, 3047 (1938).

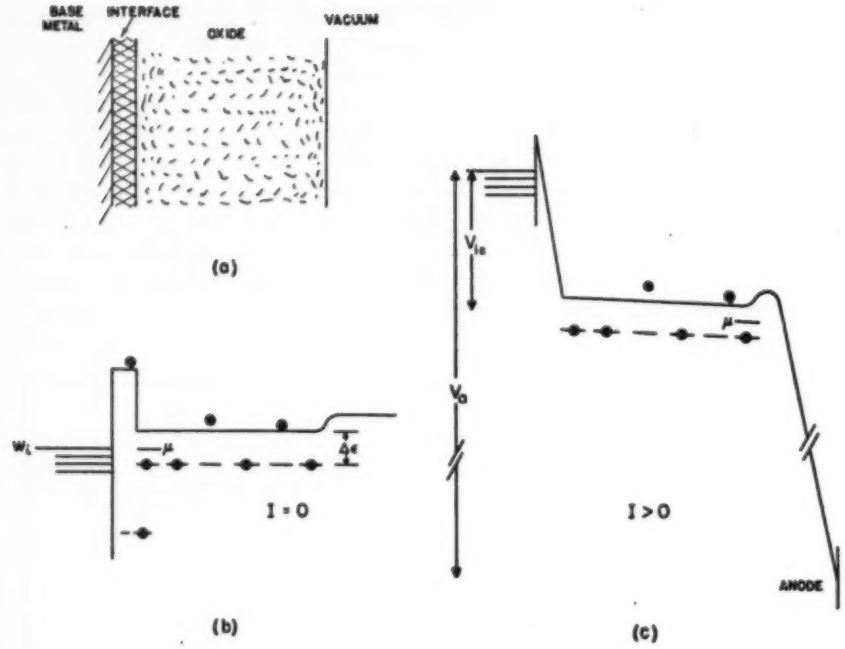


FIG. 12 (a) Cross-section view of interface type cathode, (b) energy-level representation of complete cathode for zero emission and (c) for an appreciable emission current.

Beneficial effects of H_2 on the activity of cathodes is described[25] by several workers. As in the case of high d.c. current activation, the slope of the curve decreased with increasing activity yielding activation energies between 3.37 ev and 0.53 ev for CO_2 and CH_4 treatments, respectively.

A consideration of Eq. (3) indicates that although a change in the free barium content of the sample, n_b, would shift the conductivity curve in the appropriate direction, this change alone should not alter the slope of the curve. Likewise, Eq. (3) offers no explanation for the slight change of slope with temperature observed experimentally in several of the curves. The apparent shift of activation energy with the state of activity and with the temperature is frequently observed in semiconductors and probably results from the overly simplified model of the semiconductor used in the derivation of Eq. (3). Nijboer[26] and Herring[27] have offered theoretical explanations for the change of slope over different temperature ranges; these are based on a model in which the number of impurity centers exceeds the number of free electrons. Only qualitative explanations are available to explain the decreasing slope of the curve with increasing impurity concentration, n_b. Mott[26] suggests that as the concentration increases a mutual interaction of the electrons bound at impurity centers may result, thus causing a reduction in the binding energy and the value of $\Delta\epsilon$. Any model in which a large number of electron traps are located below the impurity levels and are gradually filled as the number of impurity atoms increases, is also capable of justifying the change of slope with concentration.††††

The specific conductivity of synthetically prepared Ba_2SiO_4 was measured over the same temperature range using similar techniques, although a lower conduction current density was used because of the very low conductivity of the material. Immediately following the sample outgassing in vacuum, CO_2 was admitted and a heat treatment carried out to remove the carbon residue from the organic binder. Curve 4, Fig. 11, was taken with the sample again in vacuum and heated by the furnace. A continued heat treatment in vacuum for several hours at 1175°K produced some activation, curve 5. Passing a high d.c. current through the sample resulted in an additional activation, curve 6. This current was just sufficient to cause a noticeable increase in the temperature of the sample. A slight decrease in

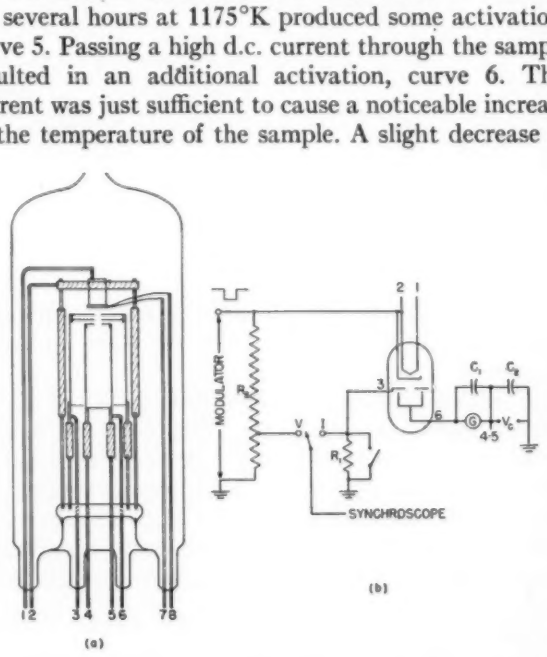


FIG. 13. (a) Cross-section view of retarding-potential tube, (b) pulsed circuit used in making retarding-potential measurements.

[26] B. R. A. Nijboer, Proc. Phys. Soc. London **51**, 575 (1939).
[27] Conyers Herring, Phys. Rev. **73**, 1238 (1948).
†††† Suggested by J. A. Burton, Bell Telephone Laboratories.

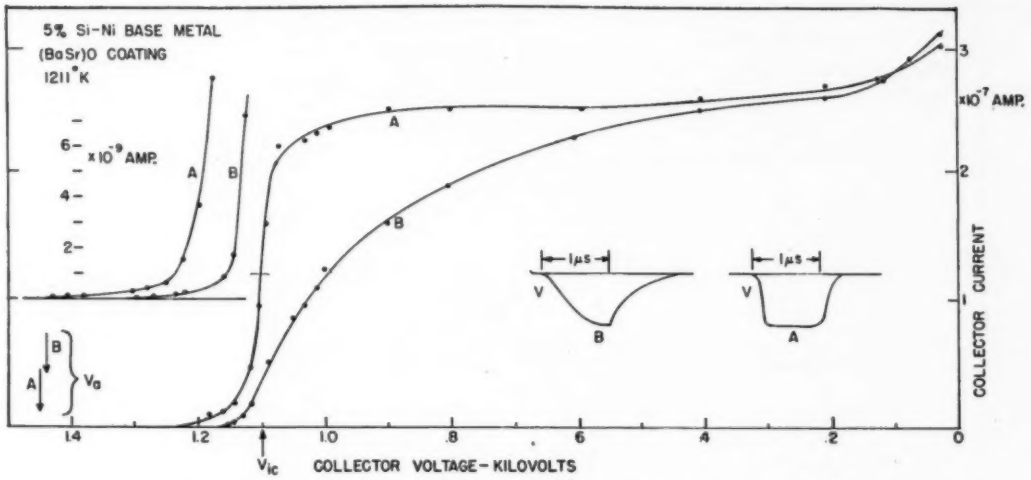


FIG. 14. Retarding-potential characteristics obtained with two differently shaped pulses.

conductivity followed a second heat treatment in CO_2, curve 7. In general, the conductivity behavior of the silicate resembles that of the oxide except that the conductivity values are much lower, and the activation energy is much higher, between 5.2 ev and 7.9 ev. For comparison, the conductivity results from Fig. 10a are shown on the same plot. It is interesting to note that at a temperature of 1250°K the conductivity of the oxide exceeds that of the silicate by a factor of only 10, while at 1000°K this factor is nearly 10,000. Any interface phenomena depending upon a voltage drop across this layer should manifest itself particularly at lower temperatures. The slight "break" in the conductivity curves suggests the possibility of two sources of electrons and two activation energies. Only the higher energy slopes are indicated by the $\Delta\epsilon$ values shown on the figure.

The presence of a low conductivity, high activation energy, interface layer must be considered in an energy-level representation of the complete cathode. One model which can be assumed is shown in Fig. 12. At (a), a cross-section view of the cathode places a relatively thick interface layer between the base metal and the oxide coating; (b) and (c) show energy levels present in this structure for the condition of zero current flow and for an appreciable current flow, respectively. Both the interface and coating are assumed to be N-type, impurity semiconductors, the activation energy of the interface being considerably larger than that of the coating. In this model the chemical potentials of the interface and coating, μ, are equated to the Fermi level in the metal, W_i. This gives rise to an energy barrier for electron passage from the base metal into the conduction band of the interface. The height of this barrier is approximately $\Delta\epsilon'/2$ with $\Delta\epsilon'$ the thermal activation energy of the interface material. Values ob-

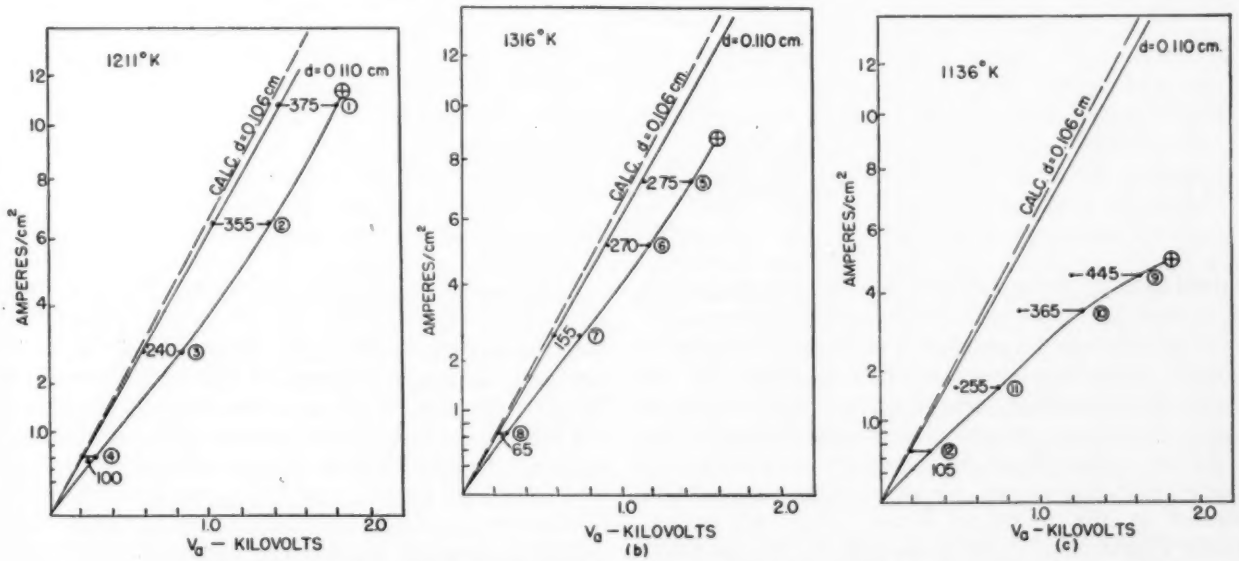


FIG. 15. Emission characteristics of retarding-potential tube at three operating temperatures.

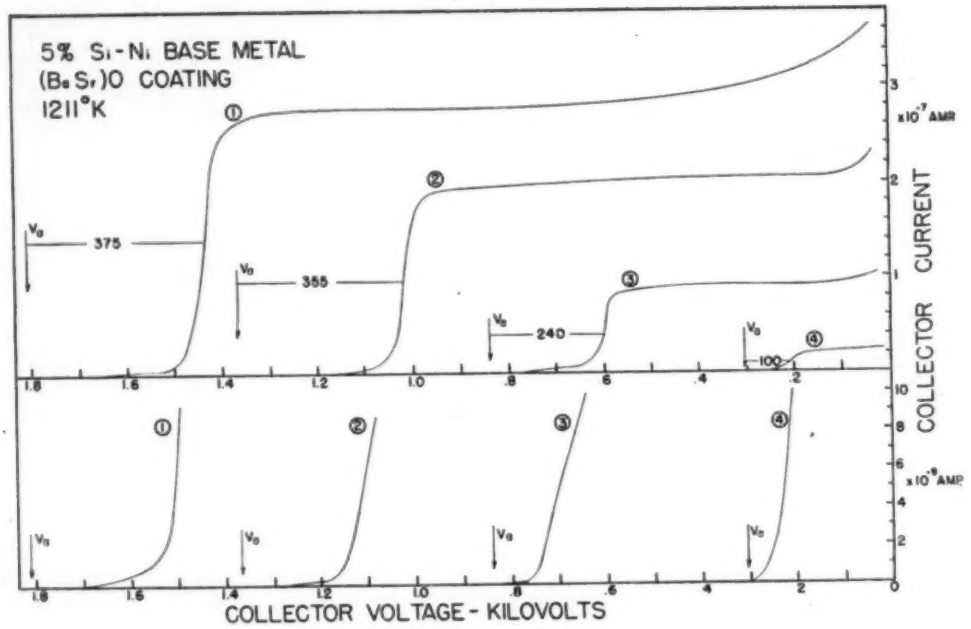


FIG. 16. (a).

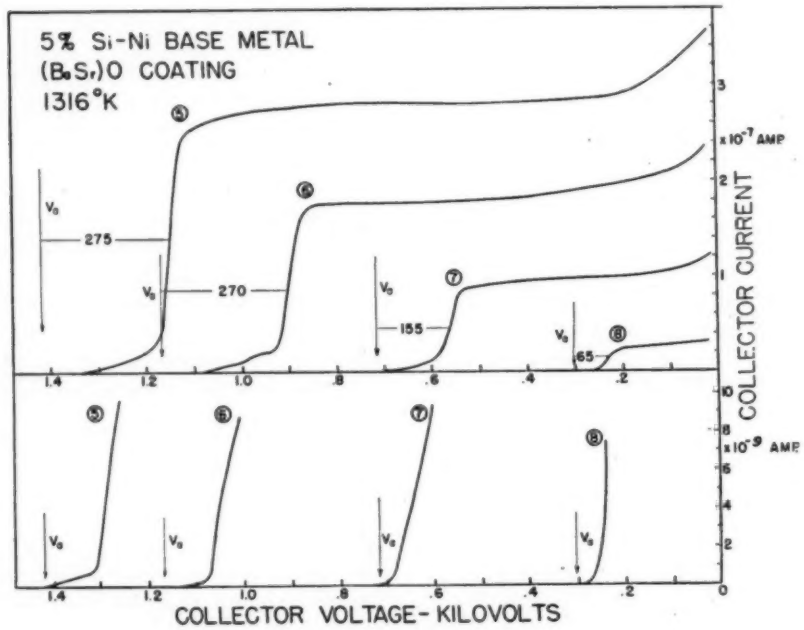


FIG. 16. (b).

tained from the slopes in Fig. 11 indicate a barrier height of at least 2.5 ev. In view of the copious electron emission current densities obtainable from oxide cathodes at normal operating temperatures, this value for the barrier height may be somewhat high. The presence of excess barium or strontium activator atoms in the complete oxide cathode may give rise to a somewhat lower activation energy for the barium orthosilicate present at the interface. Wright[7] has indicated a barrier height of 0.7 ev for a cathode whose base metal contained magnesium.

Application of an anode voltage V_a, Fig. 12c, permits a flow of electrons through the cathode which gives rise to a tipping of the conduction bands of the interface and coating. The degree of tipping is determined by the respective electrical conductivities. The chemical potential of the oxide coating near the vacuum surface is thereby depressed below the Fermi level of the base metal by an amount V_{ic}, the potential drop through the interface and coating. Under normal d.c. emission conditions this coating and interface voltage is small but under conditions of high current density,

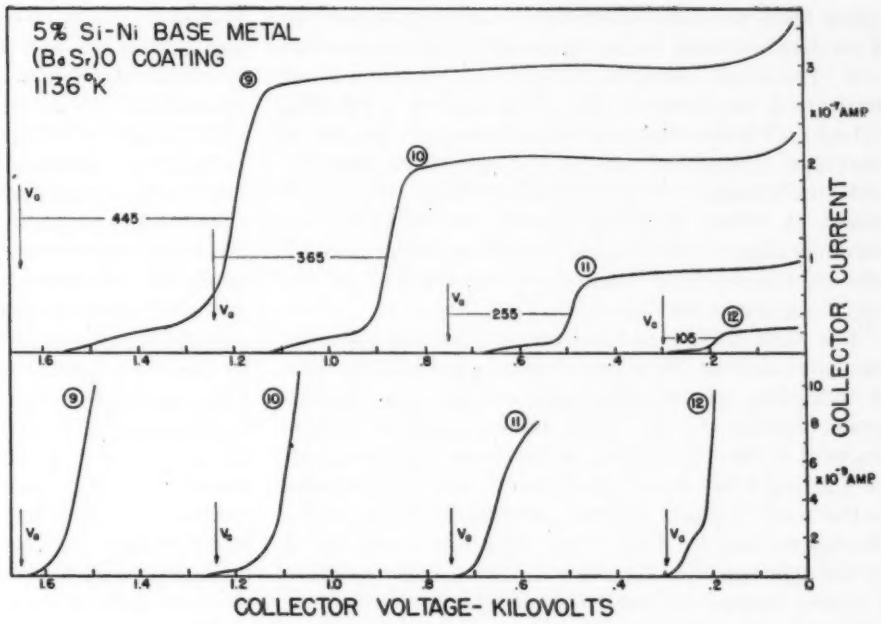


FIG. 16. Retarding-potential characteristics for three operating temperatures. (For parts (a) and (b) see page 786.)

FIG. 16. (c).

pulsed emission, or low temperature, d.c. operation this voltage is not negligible with respect to the applied voltage V_a.

IV. RETARDING-POTENTIAL MEASUREMENTS

An unambiguous determination of the oxide coating and interface conductivity may be made only by using the double-probe technique. This method[8,28,29] involves a measurement of the potentials developed in the cathode by two probes imbedded at different depths in the oxide coating. From a knowledge of the position of the probes within the cathode and the potentials assumed by each when an emission current is allowed to flow, the coating and interface conductivities are computed. Several serious limitations are imposed on this method. The construction of probe cathodes with accurately known probe positions is quite difficult. Cathodes raised to the sparking point frequently suffer a "burn-out" of the probe lead. Microsecond pulsed measurements cannot be extended to temperatures below about 1050°K because of the effect of time constants in the viewing circuit. Inherent capacity in the synchroscope combined with the rapidly increasing cathode resistance with decreasing temperature gives an *RC* time constant considerably in excess of one microsecond.

A retarding-potential method has been devised to measure the total interface and coating potential drop. Although this method does not yield values of the respective conductivities directly, these may be estimated from previous results obtained with other techniques. For temperatures below 1100°C, Mutter's[28] double-probe data indicate that the coating resistance is probably negligible compared with the interface (barium orthosilicate) resistance. This is confirmed in a comparison of the coating and interface specific conductivities, Fig. 11. The retarding potential method is not limited in its application and may be used at low temperatures as well as in the sparking region.

Electrons emitted with thermal energies from the surface of the cathode in Fig. 12c will arrive at the anode with an energy corresponding to $(V_a - V_{ic})$. These electrons are allowed to pass through a small hole in the anode and their energy is measured by observing the retarding potential which must be applied to a collector placed behind the anode to just stop the electrons. The difference between this measured voltage and the applied anode-cathode voltage is then V_{ic}.

Retarding potential tubes of the type shown in Fig. 13a were used. Flat, indirectly heated cathodes‡ having an area of 0.5 cm² were vacuum-fired at 1225°K before application of the cathode coating. A 5 percent Si-Ni base metal was used and after vacuum-firing was coated with $(BaSr)CO_3$ to a weight of 10 mg/cm². The cathode, mounted with the emitting surface facing downward, was supported from the press of the tube on insulating glass beads. The heater and cathode leads, 1 and 2, emerged from a two-lead press at the base. A 0.002-in. molybdenum-nickel thermocouple, attached to the base metal, was taken from the tube on leads 7 and 8. In operation this press acted as the cold junction of the thermocouple and was immersed in a water bath. The anode and collector were each supported on glass beads from the central press and connecting leads

[28] W. E. Mutter, Phys. Rev. **72**, 531 (1947).
[29] W. E. Danforth and D. L. Goldwater, J App. Phys. **20**, 163 (1949).

‡ Cathode parts of the type used in the 2C39 lighthouse tube were kindly supplied by H. D. Doolittle, Machlett Laboratories Incorporated.

taken from the tube on presses 3 and 6, respectively. Tantalum was used in the anode and collector construction. The anode consisted of two disks having a 2-mm separation, mounted at the upper end of a shielding cylinder. A 1-mm diameter hole allowed the passage of electrons through a nearly field-free space into the collector cylinder mounted axially inside the anode shield. A 2-mm diameter hole in the collector was carefully aligned with the anode hole to prevent secondary electrons arising from the bombardment of the upper surface of the collector.

The pulse circuit used in making retarding potential measurements on these tubes may be seen in Fig. 13b. A negative, one-microsecond‡‡ pulse was applied across resistor R_2 on which taps allowed a definite fraction of the pulse voltage to be viewed and measured on a Model 4 Synchroscope. The flow of a current pulse to the anode (3) gave rise to a potential drop across the viewing resistor R_1. This voltage was likewise measured on the synchroscope. Negative d.c. retarding potentials V_c were applied to the collector (6) through a galvanometer of sensitivity 10^{-10} amp./mm. Condensers C_1 and C_2, each 0.1 μf, provided an integrating circuit allowing the average pulsed, collector current to be read on the galvanometer. Support leads 4 and 5 were connected in the circuit as shown to minimize leakage across the glass support leads. Resistor R_1 was shorted out when retarding-potential measurements were taken.

The procedure for evaluating the coating and interface drop V_{ic} is as follows. The cathode temperature and applied anode voltage V_a were adjusted to provide operation at a given point on the voltage-current characteristic plot. As the d.c. collector potential was made negative with respect to the anode, the collector current was measured. This variation of collector current with voltage comprised the retarding potential curve and is shown in Fig. 14. The shape of this curve is closely related to the shape of the voltage pulse applied to the tube. Two pulse shapes were used and are sketched in the lower right: A is the normal 1-μsec. negative voltage pulse and B is a modified pulse, formed by placing a small condenser across the output of the pulse modulator. As the potential of the collector becomes equal to, or more negative than, the potential of the emitting oxide surface, the collector current decreases rapidly. A plot of the logarithm of the collector current *vs.* collector voltage does not yield a linear curve of slope e/kT as it should in a simple diode retarding-potential measurement. This is due to a variation in energy of the electrons arriving at the collector resulting from the changing anode voltage during the pulse. A comparison of the retarding potential curves A and B shows this effect. When the maximum applied anode voltage V_a was 1.45 kv, see Fig. 14, voltage fluctuations on the top of pulse A amounted to about ± 15 volts. This value is sufficient to explain the rate of current decrease with retarding voltage in curve A. The collector current does not drop sharply to zero, but has a "tail" of high energy electrons extending out to V_a. An expanded current scale plot at the upper left of Fig. 14 shows this tail. In many cases the tail extends out to the full applied anode-cathode voltage but never beyond.

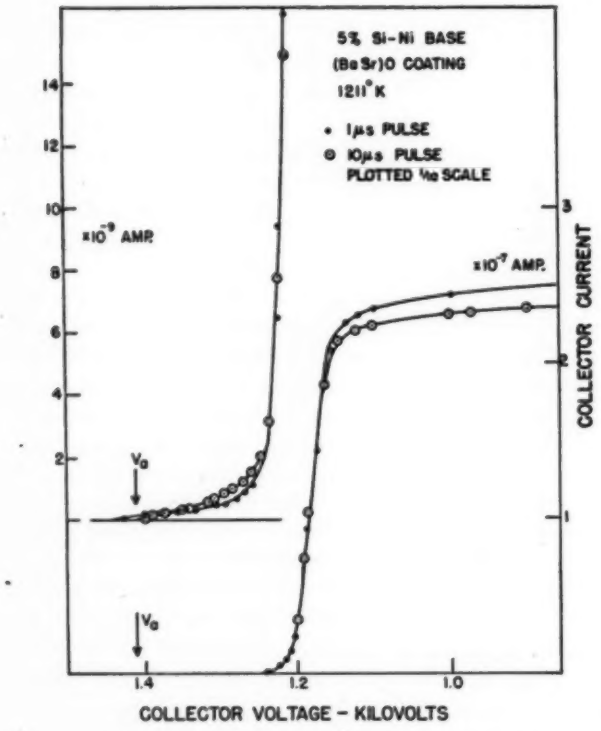


FIG. 17. Retarding-potential characteristics of 1-μsec. and 10-μsec. pulses.

‡‡ Link Model 9 Pulser.

Because of the dependence on pulse shape and the presence of the tail of high energy electrons, the usual treatment of a retarding-potential curve cannot be used. The potential of the cathode surface with respect to the anode is defined as the voltage coordinate of that point on the curve at which the collector current has decreased to half-value. This procedure defines an average measured cathode surface potential, V_m, corresponding to the average measured applied anode voltage neglecting fluctuations during the pulse, V_a. From Fig. 14, curve A, this cathode surface potential is 1.1 kv negative with respect to the anode and the coating and interface voltage

$$V_{ic} = V_a - V_m \tag{5}$$

is 350 volts. Thus it is evident that the interface and coating voltage is not negligible with respect to the applied anode-cathode voltage.

The variation of V_{ic} with cathode temperature and emission current is seen in Figs. 15, 16, and is typical of data obtained on three tubes of this type. Figure 15 a, b, c shows the variation of the anode voltage with the emission current‡ at three temperatures, 1211°K, 1316°K, and 1136°K. Experimental points 1 through 12

define the tube characteristics and the star-marked points indicate the onset of sparking. The dashed line represents the theoretical Langmuir-Child space-charge line computed for flat anode-cathode geometry and a spacing of 0.106 cm between surfaces. This value represents the average of several measurements made at different tube orientations with the use of a magnified optical projection method. Retarding-potential curves were taken at each of the twelve operating points and are seen in Fig. 16 a, b, and c. The complete retarding-potential curves are shown in the upper section of the figures with the voltage V_{ic} indicated as the difference between V_a and V_m. In the lower section, the high energy tails are shown on an expanded collector current scale. Values of V_{ic} obtained from Fig. 16 are applied to the corresponding experimental points of Fig. 15 as subtractive voltage corrections. In Fig. 15 a and b, this correction is just sufficient in each case to bring the points to the solid straight line which represents the theoretical Langmuir-Child space-charge line for an anode-cathode spacing of 0.110 cm. The difference between this and the measured spacing, 0.106 cm, is believed to represent an error in the spacing measurement since any non-uniformity in the geometry would result in a measurement which is too small. Thus it is found that the progressive deviation type of emission characteristic can be explained by the inteface and coating voltage drop.

This voltage deviation from the theoretical space-charge line cannot always be used as a measure of V_{ic} as seen from Fig. 15c. Although the subtractive correction at point 12 allows agreement with the space-charge line, points 9, 10, and 11 do not; this indicates emission-limited conditions at the cathode surface. Unless V_{ic} is known or is negligible for a given base metal and coating combination, attempts to locate the true point of deviation from the space-charge line may be hampered seriously. The use of a Schottky plot to determine the zero field emission would also be subject to error unless V_{ic} is known or is negligible.

The tail of high energy electrons, present in all cases and extending to V_a in some, is difficult to explain without making additional assumptions. Three possible causes have been considered: (1) a time variation in the potential of the cathode surface during the pulse, (2) poor electron optics allowing electrons of the same kinetic energy to be collected or rejected depending upon their position relative to the axis of the anode hole, and (3), a real variation in the energy of the electrons due to their emission from cracks and crevices below the normal surface of the oxide coating.

If the voltage drop in the interface and coating is due only to the resistivity of the cathode and the passage of current, the potential of the oxide surface should vary during the pulse only in accord with emission-current variations. Should this current lag the applied anode voltage because of capacitance effects in the cathode, this time effect could explain the presence of the tail. With the sudden application of the anode voltage, the cathode surface might rise to the potential V_a and then approach the potential $V_a - V_{ic}$ as the flow of current increases. Electrons emitted at the beginning of the pulse would have kinetic energy corresponding to V_a while those emitted later would have that corresponding to $V_a - V_{ic}$. To test this hypothesis, retarding-potential curves were taken with 1-μsec. and 10-μsec. pulses under conditions which were otherwise identical. If the tail were due to high energy electrons emitted at the beginning of the pulse, the ratio of the tail to the total collector current should vary as the pulse length is changed. Figure 17 shows the results of this test. The 10-μsec. curve is plotted on a 1/10 scale and shows reasonable agreement with the 1-μsec. curve in the regions of both the tail and the total collector current. Any change in the ratio of these currents is much less than ten to one as predicted by this hypothesis.

Poor electron optics resulting in a non-uniform electric field in the region of the anode hole could cause a variation in the effect of retarding collector potentials on the electrons, depending upon their position and direction of passage through the hole. For electrons of kinetic energy corresponding to $V_a - V_{ic}$, this effect could change the shape of the retarding-potential curve near cut-off, but under no condition could it permit the arrival of $V_a - V_{ic}$ energy electrons at a collector potential approaching V_a.

Several factors seem consistent with the hypothesis of a real variation in energy of emitted electrons. A comparison of the tails of curves *A* and *B*, Fig. 14, shows the magnitude of the tail to be associated with the duration of time that V_a remains at its maximum value consistent with this hypothesis. A microscopic examination of oxide cathodes reveals numerous cracks and crevices in the coating resulting from a decrease in density upon conversion from the carbonate to the oxide. This is true particularly for coating weights as heavy as those used in these tubes. When the cathode is heated to operating temperature, these cracks in the coating are easily visible. Cathodes prone to frequent sparking acquire pits in the coating which in some cases extend through the coating and interface. From Fig. 14 the ratios of tail current near V_a and near the cut-off to the total collector current are about 1:1000 and 1:10, respectively. It seems reasonable to expect that 1/1000 of the coating area consists of cracks extending to the base metal and that 1/10 of the area is deeply pitted.

Although previous d.c. retarding-potential measurements[30,31] are consistent with variations only in thermal energy, these measurements were made on cathodes which probably contained no interface and at currents sufficiently low that the coating voltage drop was negligible.

[30] W. Heinze and W. Hass, Zeits. f. Techn. Physik. **19**, 166 (1938).
[31] H. Y. Fan, J. App. Phys. **14**, 552 (1943).

V. CATHODE SPARKING

At normal operating temperatures the electron emission from oxide cathodes is frequently limited only by the phenomena of sparking. Although a small physical loss of coating accompanies each spark, this process is generally not injurious to the cathode emission. Many cathodes appear to spark under space-charge-limited conditions,[20] although a slight deviation from the true space-charge line due to a small interface and coating voltage may be undetected in these measurements. When the tube is operated in the region of sparking, measurements of V_{ic} should yield information regarding the effect of the interface on this phenomena particularly when the coating voltage is negligible.

A comparison of points 1, 5, and 9 of Fig. 15 shows that sparking does not occur at a fixed value of V_{ic} as the temperature is changed. Figure 5b and Fig. 7 show variation of the sparking point at several temperatures and is a general characteristic of all Ba_2SiO_4 interface cathodes. If it is assumed that the voltage deviation from the theoretical space-charge line represents only V_{ic}, i.e., space-charge limited operation, a correlation between sparking and the i^2R dissipation in the interface and coating is possible. The dash curve of Fig. 7a represents the products $iV_{ic}=900$ watts/cm^2 and offers good agreement with the variation of the sparking points. A similar treatment of the sparking points for this cathode, Fig. 7b, shows best agreement with $iV_{ic}=4800$ watts/cm^2. The three sparking points in Fig. 15 cover too small a temperature range to justify curve fitting; however, a computation of iV_{ic} yields values between 2200 and 4200 watts/cm^2.

This correlation would seem to favor a sparking mechanism based on heat energy dissipation within the cathode. Since most of the cathode resistivity is located in the interface region of thickness 10^{-3} cm, a considerable local temperature rise is expected. The above correlation (in Fig. 7) is based on the assumption that deviations from the space-charge line represent only values of V_{ic}. Points 9, 10, and 11, Fig. 15c, show that this assumption may lead to erroneous conclusions. Similarly, the agreement in Fig. 7a between the sparking point at 840°K and the dashed curve is doubtless fortuitous since this implies a 7.8 kv value for V_{ic}.

Wright[7] has assumed that the potential drop in the emitting oxide cathode occurs at the interface and has estimated the interface thickness to be such that at sparking a voltage gradient of 10^6 volts/cm exists in this region. An assumed interface thickness of 5×10^{-4} cm and the observed value of V_{ic} permit a check on this estimate. Point 9, Fig. 15c, near sparking, would lead to a value of 6×10^5 volts/cm. In other measurements, values of V_{ic} up to 850 volts have been observed, leading to gradients of 1.2×10^6 volts/cm. In view of the probable existence of thin spots in the interface layer through which the voltage gradient is in excess of these average values, the fundamental mechanism of sparking may be a simple dielectric breakdown.[29] Sufficient information is not available at the present to decide between these two possible mechanisms. The presence of a definite interface compound of low conductivity is certainly responsible for sparking and accounts for the difference in sparking currents observed in Fig. 5a, b.

ACKNOWLEDGMENT

The author expresses his appreciation to the Frank B. Jewett Fellowship Committee and to the Research Laboratory of Electronics whose joint support made this work possible.

Fields from Fluid Flow Mappers

A. D. MOORE

Department of Electrical Engineering, College of Engineering, University of Michigan, Ann Arbor, Michigan

(Received March 7, 1949)

Fluid flow within the streamline (viscous) range, made visible by appropriate means, may be made to simulate fields occurring in electrostatic, magnetic, electric current, heat flow, chemical diffusion, and other situations. Simple, reliable, and inexpensive techniques for setting up the desired fluid flow situations have been worked out, and are described herein.

Two-dimensional fields with one, several, or many non-distributed sources (or sinks) may usually be made up and operated with ease. Techniques for the simulation of certain three-dimensional fields with axes of symmetry have also been developed.

Through the invention of the "sandbed" feature, the fluid flow method has been greatly extended: fields due to distributed sources can be simulated, not only outside the sources, but inside them as well.

Numerous photographs of fluid mappers in operation, illustrating all of the foregoing types of cases, are included.

1. INTRODUCTION

IN fluid mechanics, very broad use has been made of fluid flow made visible, for solving the problems of fluid flow itself. However, the use of streamline flow for simulating other phenomena, and solving their problems, has seen little development. Nearly half a century ago, some beautiful pioneering work was done in England,[1] making fluid flow portray magnetic fields.

[1] Hele-Shaw, Hay, and Powell, "Hydrodynamical and electromagnetic investigations regarding the magnetic-flux distribution in toothed-core armatures," J. Inst. Elec. Eng. 34, 21–53 (1904–5).

Why has this method—with its great potential value for teaching and problem-solving—remained undeveloped through all these years? From the writer's recent experience in perfecting the techniques described herein, it is suspected that lack of simple and reliable techniques has been the main handicap.

In seeking a way to get a visual simulation of field phenomena concerned with a distributed source, the writer invented the sandbed mapper in 1943. A model was at once built and successfully operated, first with air as the fluid, and then with water. Only recently has time permitted the fuller development of the sandbed mapper. Fortunately, the same techniques needed for a sandbed mapper are, aside from the sandbed feature, applicable to any number of simpler fluid mappers.

Many of the flow patterns are of great beauty, appealing to the layman as well as the scientist or engineer (see Figs. 11, 13, 30). In fact, one who sets up a fluid mapper laboratory will have to resign himself to running a dynamic art gallery in addition to handling the technical interests.

The writer has concentrated for many years on graphical field mapping, and probably has had the privilege of teaching graphical techniques to more students, than any other. The numerous fluid mapper cases operated for classes during the present school year have amply demonstrated that a student's speed and comprehension in grasping the nature of a field are greatly accelerated by these devices. The student may sketch directly from the mapper; or again, the flow pattern may be photographed, enlarged, duplicated, and handed to the student. This does not mean that he is handed a complete solution of his problem: plenty of work remains to be done to put in the orthogonal system, and to subdivide the map so that it is ready for making numerical solutions.

2. BASIC ELEMENTS OF FLUID MAPPERS

The components of a fluid flow mapper are a slab, such as any slab in Fig. 1; a piece of $\frac{1}{4}$-in. plate glass, larger than the slab, parallel to the slab surface, spaced upward from the slab by a constant distance—thus creating a two-dimensional flow space; a tray for immersing the combination in water, Fig. 3; one or more tanks with any desired water levels, Fig. 3; rubber connector tubes ($\frac{1}{8}$-in. I.D. with $\frac{1}{16}$-in. wall), each connecting a tank to the sealed bottom of a slab hole, or source; crystals of potassium permanganate sprinkled on the slab prior to placing the plate, to make the flow visible; and barriers within the flow space, if called for. A sandbed component will be described later.

3. THE SLAB

The problem of easily creating a flat uniform flow space was solved by casting a plaster slab against plate glass. At first, patching plaster, mixed with white sand, was used. Better results came when the sand was omitted. Very good slabs can be made of patching plaster only. The next and final step was to learn how to use what the dentists call artificial stone. Actually a plaster also, it is a high quality, uniform product, with a setting expansion of the order of 0.1 percent. It comes as a powder, in ivory or white. Either color will do. The writer prefers a white slab, with additions (such as bottom seals and feet) made of ivory. The powder must be added to the water, and not the other way around.

By now, the writer has used well over 100 pounds of powder. Powders from four different firms (Albastone, Castone, Rapid Stone, Crestone) have been tried, and all are satisfactory. However, the formula given for dental use is too dry for slab-making. Five pints of powder to two of water is best for slabs, and these actual volumes are used for the square slabs shown herein. These are one foot square and about $\frac{7}{8}$ in. thick. Such a slab will cost a dollar or less, for powder. The cost for patching plaster would be about half that, but the saving is not recommended.

For permanent slabs, or high accuracy slabs, cast aluminum may be the best material. For use with separate barriers, the top would be milled flat; or to dispense with separate barriers, the desired flow space would be milled into the slab face.

4. CASTING THE SLAB

In Fig. 2, a number of techniques for casting the slab used in Figs. 14–17, 25, and 26, are shown. The slab is cast on the drawing board, literally speaking. This slab is to have four little holes in a row, and one large rectangular hole, any one of which may become a source or a sink. On the drawing board is a drawing, showing the outlines of the slab, the rectangular metal core, and the four brass tubes to be cast in the slab and left as inserts. The tubes, now bottoms up, already have caps soldered to seal the bottoms, and side-tubes soldered in to take the rubber connectors. Resting on the drawing is the 14-in. square plate (plate glass). Resting on it are the metal parts, located by eye above their outlines. Also resting on the plate is a brass "fence" or outer mold. It is of brass strip, 1 by $\frac{1}{16}$ in. in section. Two opposite corners of the fence have had corner pieces added on and soldered. The other two corners come apart later. Their (loose) corner pieces may be temporarily held on with paper clamps, until a plaster slush is applied.

Unless metal parts are anchored, pouring the mix would always slide them out of place along the smooth plate. A small batch of plaster is mixed, and applied as a slush with a spoon, all around the outside of the fence, letting it pile up against the fence and also run out on the plate somewhat. First, on setting (in 10 minutes or so) it firmly anchors the fence. Second, it acts as a seal, preventing seepage of water from slab mix out under the fence—thereby giving a better edge to the slab. Slush is also laid down inside the metal core in like fashion. Any such core as this had better be slightly tapered. As to anchoring the metal tube in-

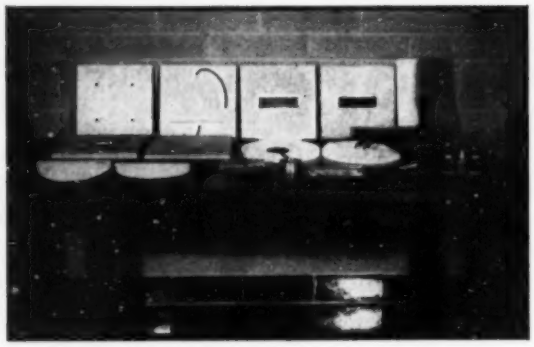


FIG. 1. Slabs of artificial stone, and some accessories.

serts, plaster techniques again furnish the easy way. In Fig. 2, a light metal bar has been laid on top of the tubes and plastered to them. This bar extends over the fence, and metal blocks standing on end are plastered to the bar, and to the fence and plate. Anchorage plaster

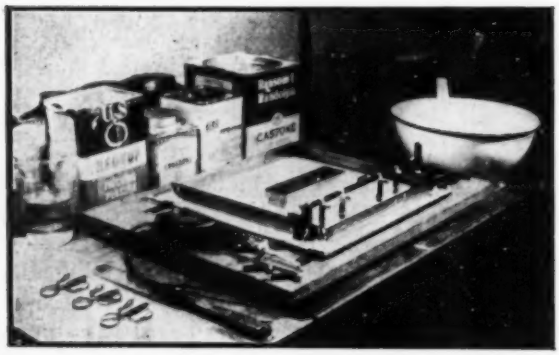


FIG. 2. Preparation for slab-casting. The brass containing fence or mold rests on plate glass, which rests on the drawing on drawing board. Central metal core is in place, as are four metal tube inserts.

is easily cracked off after the slab pouring. Metal parts in contact with the slab itself are painted with a thick soap solution to prevent sticking; or Silicone DC4 Compound, which works better than soap, may be used.

All the water is put in the bowl, Fig. 2. The powder, already measured, is poured in while rapidly stirring.

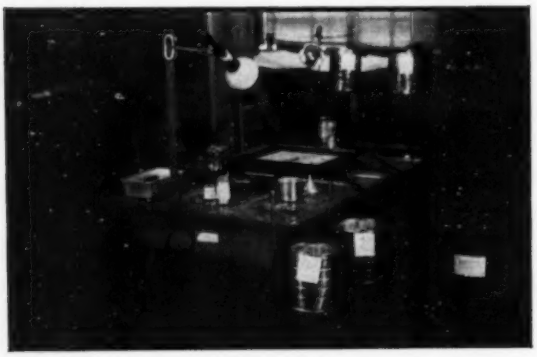


FIG. 3. Fluid mapper in operation. No. 2 photofloods are in place, for photography.

Stirring will take about a minute. Some small lumps may then remain, but these will be fully wet, and they will cause no difficulty. Mechanical mixing would be desirable. The resulting mix is soup, like pancake batter. It can be poured with little care; it is spread around with the paddle, and is evened up by jiggling with the paddle.

Within a few minutes after setting, the anchorage plaster is cracked off, the fence is removed, and the plate and all that it bears is set in water. The plate is then slipped off with ease. The metal core is tapped out; sometimes, it can simply be lifted out.

While the slab is still soft, a sloping channel is cut into the bottom, running into the side of the central hole. It is to take a rubber connector tube. A thin aluminum plate, about $\frac{1}{4}$ in. larger all around than the hole it is to cover, is placed over the bottom of the hole. It and the rubber tube running in under it are plastered to the slab by additional mix, which is slushed around and over the edges of the plate. At this time, low feet are also plastered on by piling up plaster where desired. The slab must be completely wet—but not dripping—when any such plaster is added. Any such additions hold with ample strength, but may easily be cracked off at any time with a chisel, to make repairs or changes.

5. OTHER SLAB TECHNIQUES

One thinks of the hard ways first—and this applies to the above method of getting four little sources by means of metal inserts. Someone may sometime need metal inserts, and therefore, the details were described. However, such inserts are most unwelcome if the slab surface becomes imperfect from use, and needs to be ground flat again. Also, small holes are more easily made by drilling, from the top down, with bottom seal and rubber connector to be plastered on. For $\frac{1}{4}$-in. round holes, the rubber connector is wetted with plaster and stuck directly into the hole bottom. Large holes are best made with hollow metal cores, internally anchored with plaster.

Artificial stone offers many advantages: it may be sawed, carved, routed, drilled, reamed, end-milled, and so on, after firmly setting, either when new and still damp, or when air-dry. When dry, it can be ground or sanded.

The one bad quality of the stone is its erosion in use. The mirror finish it has when first cast, rapidly disappears when immersed. The slight roughness resulting is of no importance whatever, but the erosion is. In several hours of use, the areas of higher fluid velocities may erode as much as 8 or 10 mils. Many of the photographs shown herein were taken before erosion was recognized, and such patterns must be taken as good demonstrations rather than accurate simulations. An eroded slab can be retrieved by first letting it dry, then grinding it by turning it face down and moving it against emery cloth. Flatness is tested for by slipping feeler gauges under plate glass applied to the slab sur-

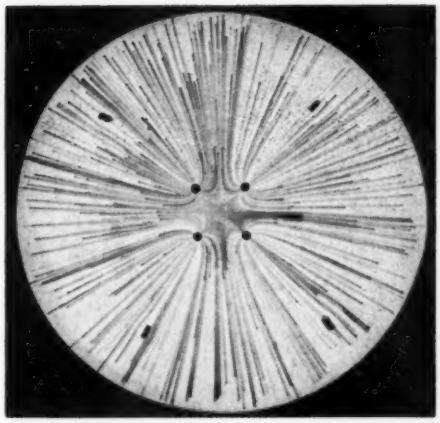


FIG. 4. Four equal sinks. Each streamline begins at a crystal of potassium permanganate. The plate glass or plate, spaced $\frac{1}{16}$ in. above slab, does not show. It is covered with water.

face. Some of the writer's experience indicates that once the first few mils are thus removed, erosion may proceed more slowly. Often, it will be easier to throw the slab out and build a new one. Prevention of erosion by finding a suitable surface treatment is due for further study.

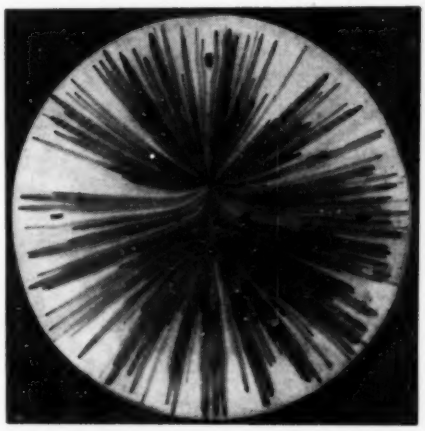


FIG. 5. Two unequal sinks, approximately 2:1 in strength.

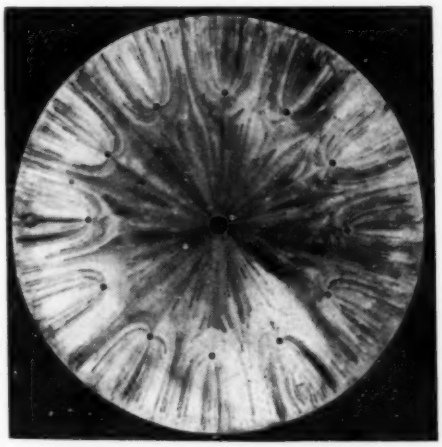


FIG. 6. Simulating a triode's field. Central cathode, 12 axial grid wires. All holes operated as sinks.

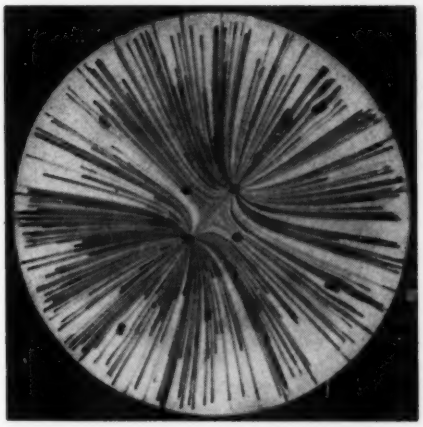


FIG. 7. Two equal sinks.

6. SPACERS, BARRIERS, SEALANTS

In all but one (Fig. 29) of the two-dimensional cases shown herein, a flow spacing of about $\frac{1}{16}$ in., or about 60 mils, was used. Spacers are cut from sheet rubber or from brass strip.

The best material for general barrier purposes so far

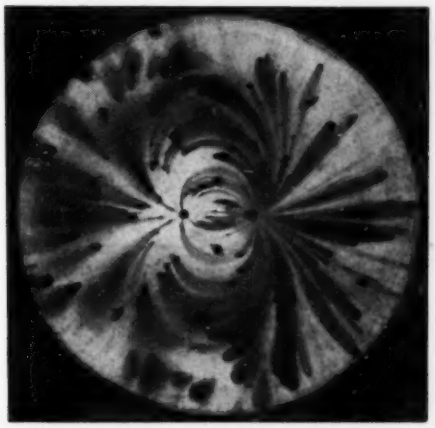


FIG. 8. Source and sink, equal.

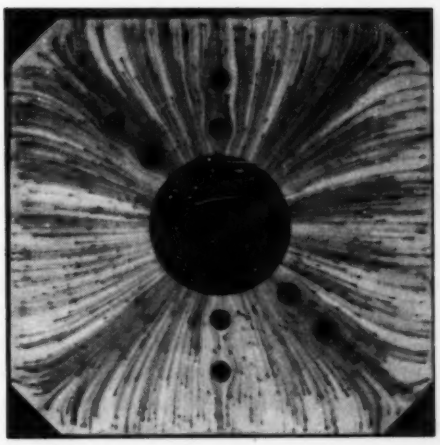


FIG. 9. Simulation of a heat flow case. Round black rubber pieces represent thermocouple wells.

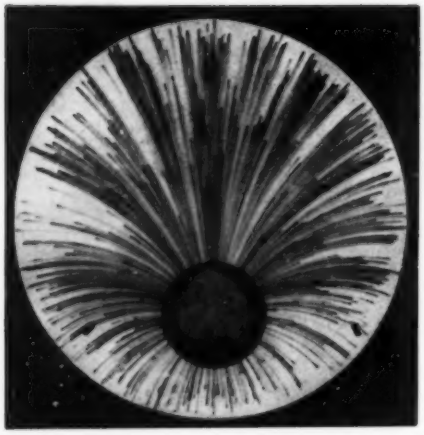


FIG. 10. Open-edge slab, large round sink eccentrically located. Screen for supporting a sandbed is seen in the hole.

FIG. 11. Square slab, open edge. Four small equal sinks, large round central source.

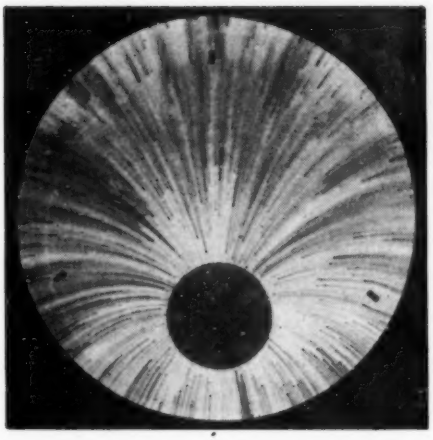


FIG. 12. Sandbed operating as a distributed source. Note displacement of kernel upward from source center.

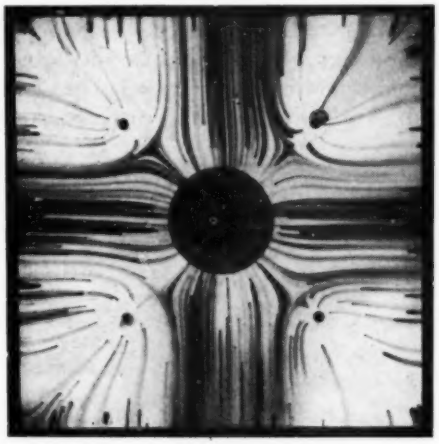


FIG. 13. Central sandbed as a distributed source, and four other small equal sources.

found is $\frac{1}{16}$ red sheet rubber; even though not finished smooth, there usually has been no leakage on account of the finish. At one time, some smooth black sheet rubber, with cloth insert in the middle, had to be used. It was so stiff that permanent set prevented it from lying flat willingly. Brass strip, flattened by hand bending and hammering, and polished, works nicely: it is heavy enough to remain located better than rubber, when water movement occurs due to placing the plate. However, for a variety of purposes not calling for high accuracy, high grade sheet rubber is most easily worked up.

Leakage across a barrier due to barrier or slab imperfections may be absent, or too slight to worry about. When bad enough, a sealant is called for. Some success has been achieved by smearing Vaseline on the barrier before placing it. It is not an invariably convenient and reliable sealant. One night, when it failed, the writer turned in desperation to a sample tube of anti-fogging material that was accidentally at hand. It worked well, and later on, did so again. This stuff, now obtainable at filling stations, is a clear, viscous something made "to prevent fogging on glass surfaces." At this writing, it is the only sealant that can be recommended. It should not be applied so that it will be squeezed into the flow space: it is nearly invisible under water, and might become an undetected block to the flow.

A search is now being made for some rubber-like sheet plastic, which, it is hoped, will be smooth-finished and highly uniform.

7. MAKING FLOW LINES VISIBLE

Much fluid flow work in applied mechanics is done with a free surface; thus, injectors, injecting colored streams of fluid such as Malachite Green, are feasible. Such a technique could seldom be easily applied to fluid mappers, where no free surface exists. Potassium permanganate crystals therefore are used. These come in mixed sizes. After crushing somewhat, they are screened to eliminate the finer particles. Those remaining should freely come through holes of about 1 mm diameter in the head of a saltshaker. Such crystals will

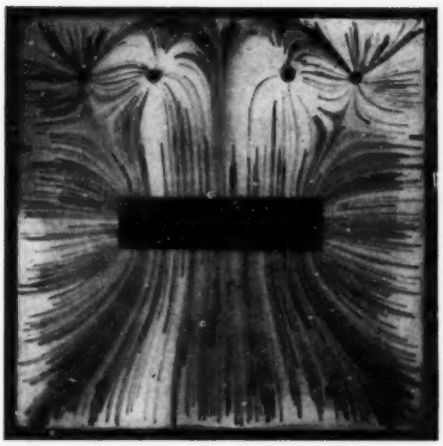


FIG. 14. Square open-edge slab; large rectangular central sink; two outer holes, equal sinks; two inner holes, equal sources.

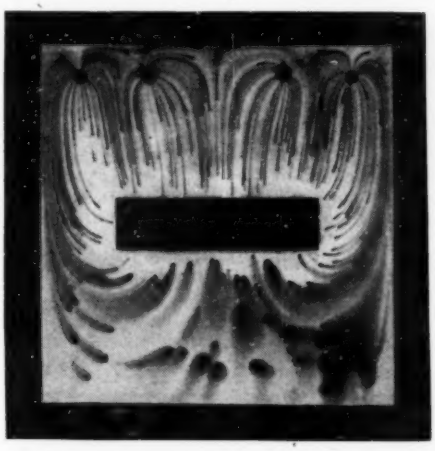


FIG. 15. Flow contained by sheet rubber barrier all around edge. Large central source, four equal sinks. In lower part where velocities are low, lines are broad due to sidewise diffusion of dye.

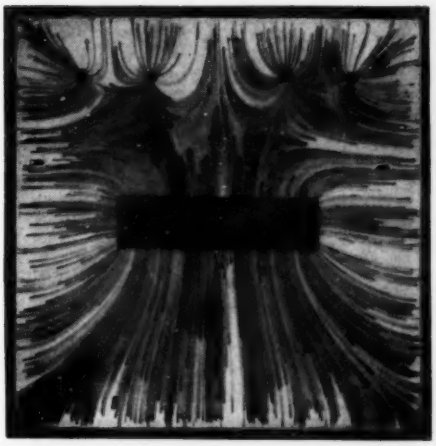


FIG. 16. Square open-edge slab; four small equal sinks, large central sink.

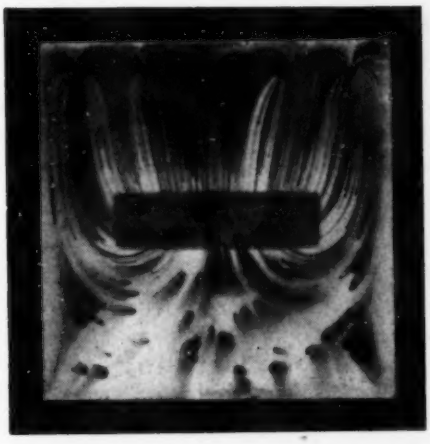


FIG. 17. As in Fig. 15, but with sandbed simulating a distributed source.

last for many minutes, each crystal forming the head of a streamline of color.

In fluid mapper use, the permanganate often changes fast enough to form a brown stain; in prolonged operation, the MnO_2 formed may appear as a precipitate, moving along with the stream flow. A stained slab can be completely cleared in a short time by brushing its face with this mixture: 95 parts of 3 percent hydrogen peroxide, 5 parts of acetic acid, 300 parts of water. Any rough approximation to the formula is effective. The slab is of course raised above water for this treatment. If immediately put into service again, brown precipitate will copiously appear: peroxide, diffused into the slab, diffuses out again to unite with the permanganate. After clearing a slab, it should be flushed with water, then allowed to soak in water for a while.

In spite of diligent inquiry among chemists, the only other material now known to the writer which might be suitable, is methylene blue. For years, the writer has had a little of this in crystal form, has used it for making streamlines in water many times, and found it better than permanganate in sharpness of line. Recently, the supply was all used up. Then—no one had any more, and no chemist yet approached knows of methylene blue being successfully made into large crystals of 2- or 3-mm size. It may be that the "crystals" one had were fake—made by boiling the dye with soluble starch, gelatin, etc., spreading the mixture out to harden, then cracking it up. It is to be hoped that some such crystal coloring agent, better than permanganate, may be found. MnO_2 precipitate is not an asset to fluid flow devices.

A very striking and often valuable technique is to color a whole area of flow. Instead of using crystals in the flow space, one of two flows, say, is colored by coloring the water in its feed tank; and the other flow is left clear. This was done in Fig. 20, the set-up being the same as in Fig. 18. The two sources at the right were supposed to be equal, and the sinks at the left, equal. The horn of colored fluid extending leftward from the inner sink may be partly due to set-up imperfection;

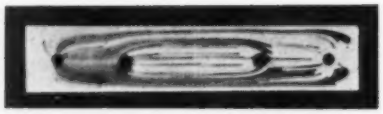


FIG. 18. Equal sources at right, equal sinks at left.

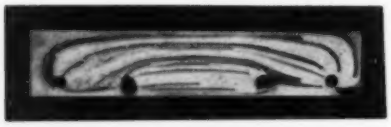


FIG. 19. As in Fig. 18, but sources and sinks displaced downward. The particles are MnO_2 precipitate.

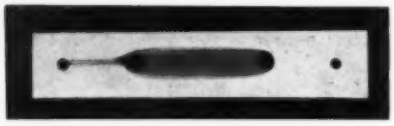


FIG. 20. Case of Fig. 18, but no crystals used. Entering fluid colored for one source, but not the other.

but also, this effect can ensue due to lateral diffusion of the dye, from the colored part to the uncolored flow area.

8. THE TRAY

A rustproof tray should be used. However, the tray of Fig. 3 is a cheap black steel kitchenware tray, with sloping sides. It is repainted from time to time. It is 3 in. deep, and about 15 by 16 in. horizontally. For convenience, several junction tubes are provided. These are thin-wall brass tubes, $\frac{5}{32}$-in. O.D., piercing the tray wall and soldered into it. Rubber connectors from the slab join to the inside projections; connectors from tanks go to the outer ones. Thus, connectors need not be brought out of the tray by coming up over the edge.

9. OPERATION

The slab is put into the tray, with enough tap water used to cover the plate when it is in place. Rubber connectors are hooked up. Water is put in the tanks to make their levels, at first, about the same as tray water level. Barriers or spacers, as the case may be, are put onto the slab surface, under water. Crystals are

FIG. 21. One source at lower right, and three sinks.

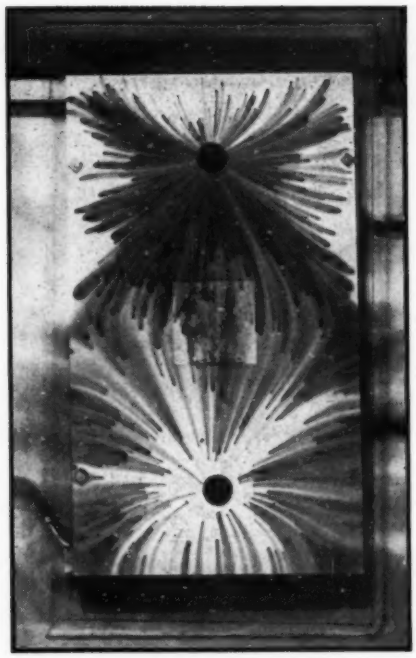


FIG. 22. Open-edge slab; source and sink equal. Paraffined bristol board embedded in slab still in place; uniform flow space. Note swirl effect on flow lines leaving the source.

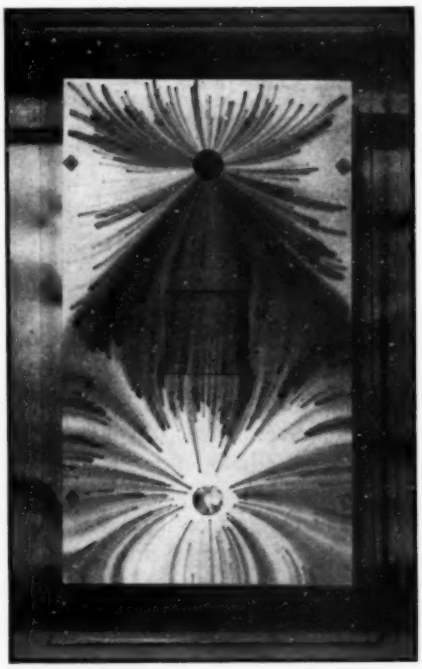


FIG. 23. Flow space increased in center square by removing embedded insert shown in Fig. 22. Note the refraction. Note also the gauze crowded into lower half of source hole, curing swirl effect.

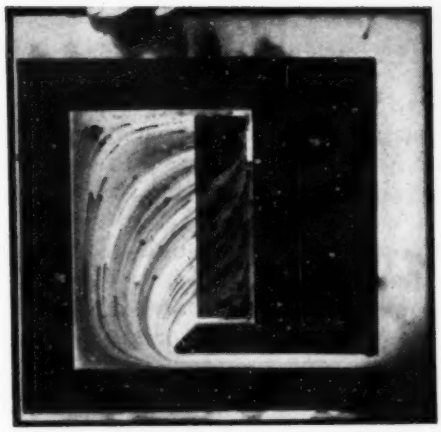


FIG. 24. Sandbed mapper simulation of equipotentials of magnetic field due to current distributed in field coil wall section.

sprinkled over the exposed slab areas; they promptly sink to the slab. The plate, preferably cleaned with Glass Wax or its equivalent, is slowly lowered. On placing the plate, water is displaced outward, and a "placing pattern" is formed. After operation is started, part of the placing pattern may remain for quite a while, in low velocity regions.

Consider Fig. 4, an open-edge case. The four holes are to become equal sinks. Equal lengths (each about 50 in.) of rubber connectors have been used. Tank levels have been made the same as tray level. All four cans are now picked up together, moved off the table, and lowered a few inches. The pattern at once begins to form. With the flow space and rubber connectors as described, most of the flow resistance will be in the connectors. Connector inequality will occur for three reasons, in spite of using equal lengths. First, rubber tubing is not uniform; and in a long round straight duct, resistance, or ratio of pressure drop to flow rate, varies inversely with diameter to the fourth power. Second, there are bend effects. Third, there are pressure drops at entries and exits. The second and third factors rise, relatively, with rise in velocity. Even so, many nearly symmetrical patterns, in symmetry cases, can promptly be obtained by simply using near-equality of equal-length connectors. When better symmetry is desired, a little water can be added or taken from one or more tanks.

When one pattern is changed into another, as by setting up Fig. 4, then clamping two connectors to get Fig. 7, the effect is indeed striking. Nearly every spectator comments on it. First, there is the old pattern; then apparent indecision; then the new pattern suddenly seems to appear everywhere.

For Fig. 5, the two tanks may be lowered by different amounts, to get operating heads of 2:1. Or, Fig. 8, equal positive and negative heads can be used, to give a source and a sink of equal strength.

Inertia effects are nil when the flow space is operated within the viscous (laminar or streamline) range. This is strikingly shown by setting up, say, the Fig. 4 pattern with half-inch heads; then suddenly lowering the tanks to the floor, to give about 30-in. heads: the pattern is unchanged. A large majority of cases can be operated with heads of 2 in. or less.

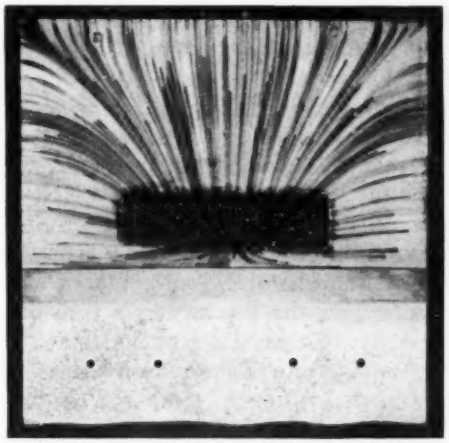


FIG. 25. Sandbed sink, brass barrier below. Crystals on slab. Others attached to plate above sandbed edge, to get good delineation over sandbed.

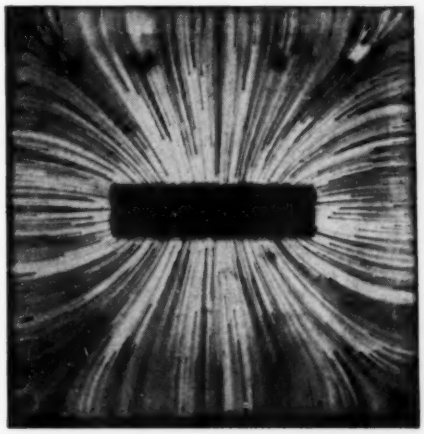


FIG. 26. Open-edge slab, central sandbed as source.

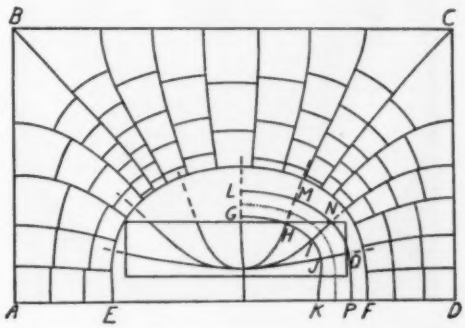


FIG. 27. Mapping study of the case shown in Fig. 25.

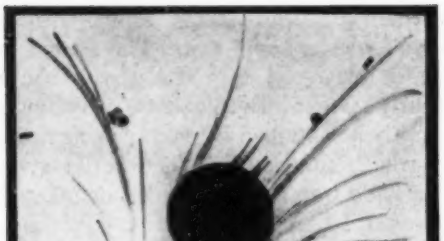


FIG. 28. One small sink, dead sandbed. Refraction at sandbed edges, and lines avoiding it, show that sandbed is high.

10. SWIRL EFFECT

In all cases not involving a sandbed, the streamlines should approach or leave the edge of the sink or source orthogonally. Likewise for an open slab edge. However, in Fig. 22, it is seen that whereas the fluid approaches the (upper) sink at right angles to the edge, it is not leaving the (lower) source properly. If the hole were large compared to the flow spacing, this trouble would not occur. As it is, water entering the tube at the bottom retains a swirling movement, and local velocities in the hole are large enough to affect nearby flow in the flow space. The cure was to stuff gauze into the lower half of the hole, to even up the velocities in the hole. In the (lower) source, Fig. 23, the gauze can be seen; and it has also been used in the two sources of Fig. 35.

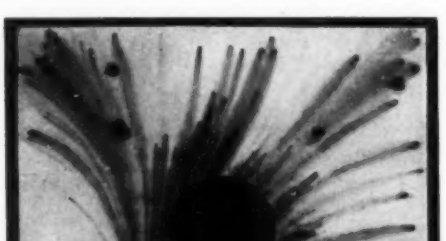


FIG. 29. Same as in Fig. 28, but flow space doubled. Lessened refraction shows that the larger the flow spacing, the less the effect of spacing inequalities.

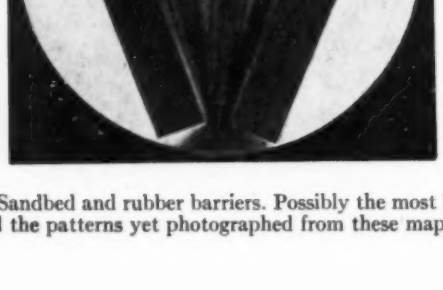


FIG. 30. Sandbed and rubber barriers. Possibly the most beautiful of all the patterns yet photographed from these mappers.

Figure 6, simulating a triode with central cathode and 12 axial grid wires, at first suffered from swirl effect, when holes were sources, distorting part of the field near the grid wires. That was when this slab looked like a Portuguese Man-of-War, with 13 rubber tubes trailing out from underneath. This was changed by removing the 12 separate connectors; beveling the 12 holes from below with the end of a large drill, to equalize hole lengths and approaches; and installing underneath a common plenum chamber. Several baffles of fine screen were mounted in front of the tube feeding the plenum chamber, in the chamber itself. Symmetry of pattern at once resulted.

11. SIMILITUDE

In Fig. 4, the open slab edge is an isopressure boundary. Calling the sinks equal, their edges are isopressure boundaries, all at the same pressure. Then the flow set up in the uniform flow space must be normal to these boundaries. Also, what we may think of as flow resistivity offered in the flow space is constant everywhere. Reinterpret it as a current flow case; current is flowing in a uniform homogeneous sheet between boundaries made into equipotentials by suitable electrodes. The conditions are such that similitude occurs between the flow pattern and the current pattern. Interpret again as a heat flow case in a sheet, when equipotentials are isothermals, and similarity again prevails. Or again, four small parallel cylinders, all at one potential, set up an electrostatic field between them and an enclosing cylinder. Or, interpret as a magnetic case, a chemical diffusion case, and so on.

Returning to the current-in-sheet interpretation, if we go vertically through the conducting sheet, constant current density is found; whereas, in the fluid flow space, a parabolic velocity distribution occurs. This departure from similitude is of no concern, for the streamline pattern set up in any one level duplicates that at another level.

Figure 9 had its origin as a heat flow case. Mr. J. L. Schweppe, Ph.D. candidate in Chemical Engineering, is preparing to do research on evaporation of liquid in a central hole in a metal bar, heat being forced into parts of the four sides of the bar. The round spacers represent thermocouple wells. In order to know the facts about the heat flow pattern and the meanings of

thermocouple readings, slabs were built and operated for three sizes of holes. To show how speedily some cases can be carried through, take this case. The hole cores were turned out in advance, and the barriers and spacers cut. Then, beginning one morning at 9:00, the writer, with help from Mr. Schweppe, cast all three slabs, operated and photographed them, and had the films developed by 5:15 P.M.

Figure 35 may be thought of as a three-phase line in an iron conduit, at the instant when one conductor has maximum current. Fluid flow lines correspond to magnetic equipotentials. The writer recently carried the materials, mold, cores, and suchlike to class; in one hour—lecturing meanwhile—this slab was cast, and the appendages added; in the next hour, it was operated.

12. MAP COMPLETION

This paper can merely mention map completion steps. From a fluid mapper, a flow pattern is taken off, photographically or otherwise. The orthogonal system can sometimes also be set up on a fluid mapper; if not, the operator must put it in himself by suitable graphical methods. In the end, the map consists of selected lines in both systems. A two-dimensional map is commonly made up of a regular array of curvilinear squares. Moore[2] may be referred to for graphical field mapping techniques. Numerical results are easily obtained, once the map is completed.

13. THE SANDBED MAPPER

The invention which was the major incentive for developing these fluid mapper techniques, was the sandbed feature. Consider Fig. 10 as a heat flow case in a long solid cylinder having an eccentric axial hole. If hole and cylinder surface are isothermals, heat flow lines would be as in Fig. 10. Next (Fig. 12) let the cylinder be without the hole, and let heat be produced constantly throughout that part of its volume where the hole was. Heat would flow away from every point where it originates, in some direction, *seeming* to originate at a point called the kernel. To simulate, the hole in the slab is provided with a screen, halfway down, visible in Fig. 10. A uniform bed of sand is filled in on the screen, with top sand leveled off just flush with the slab surface. Flow spacing, sand grain size, and sandbed depth are so selected that the horizontal flow space resistance is negligible compared to that of the sandbed. Operating as a source, fluid would have to appear uniformly above the sandbed, then flow away in the flow space. This corresponds to heat appearing uniformly within the volume described, and there will be similitude.

The kernel cannot occur at the sandbed center, but is displaced upward, toward the region of higher flow space resistance. Foreseeing the probable location of the kernel in a new case seems to present the mind with an unusually difficult puzzle. When a kernel which otherwise would be at a center of a symmetrical source is to be displaced by change of barriers or what not, students at first (and others too!) almost invariably predict a displacement opposite to what really occurs.

[2] A. D. Moore, *Fundamentals of Electrical Design* (McGraw-Hill Book Company, Inc., New York, 1927).

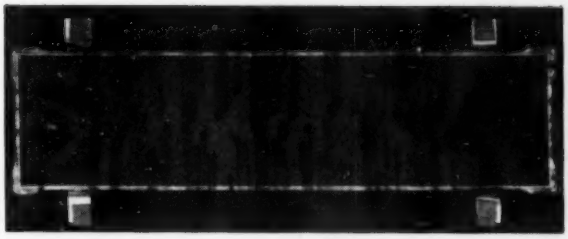


FIG. 31. Isolated sandbed, 2 by 8 in. Sand: cast iron shot copper plated. Operated as source, and crystal-sprinkled in usual way.

Re-interpreting, the sandbed becomes the equivalent of a uniform space charge; a distributed current setting up a magnetic field; a uniform and changing flux piercing a disk, setting up eddy currents both outside and inside the source; and so on. Problems in these situations typically defy mathematical solution; approximation methods can be used, but cause a great deal of work to be done. The very small literature measures the lack of progress to date. The sandbed mapper can yield a very good first approximation with simple techniques; or certainly can solve a case with high accuracy, if accurate devices are built.

FIG. 32. The case of Fig. 31 operated as a sink, crystals attached to plate by "nail polish." Better lines obtained.

The figures show a number of sandbeds in operation. Figure 24 shows (developed) half a pole pitch of a machine, the sandbed simulating the coil wall section current. The slab of Fig. 34 was built to solve for Foucault currents in a round disk. Take the radial

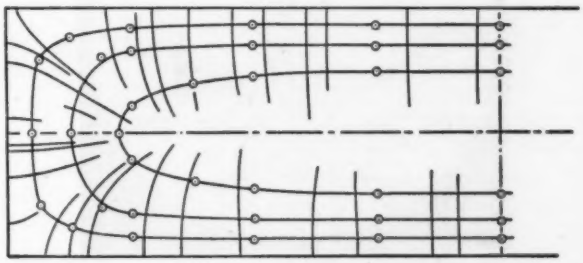


FIG. 33. Lines from the isolated sandbed, orthogonally crossed by contour lines of soap film produced on top of the empty sandbed frame.

FIG. 34. Open-edge slab, small narrow sectorial sandbed used as sink. Solving for Foucault (damping) currents in a round disk, rotating through a flux piercing the disk.

dimension of the little narrow sectorial sandbed, and sweep it clockwise to 12 o'clock. Imagine the other half to be present, with a similar area. Fill this total area with uniform flux, and rotate the disk through it. Then

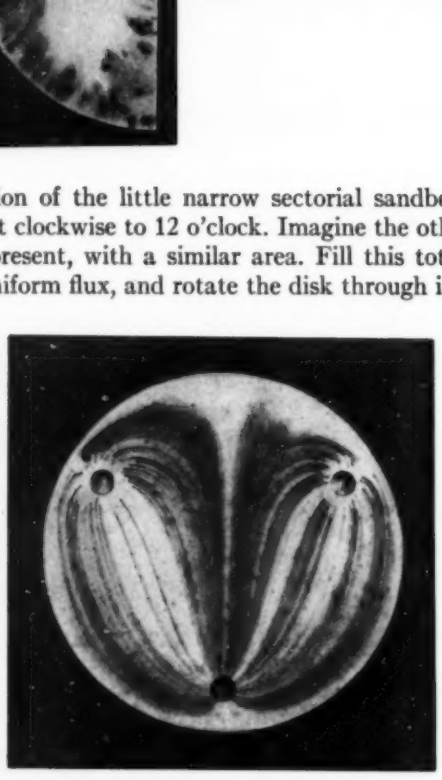


FIG. 35. All-around barrier. Two equal sources (note gauze to cure swirl effect) and a sink. Solving for magnetic field due to three-phase conductors in an iron duct.

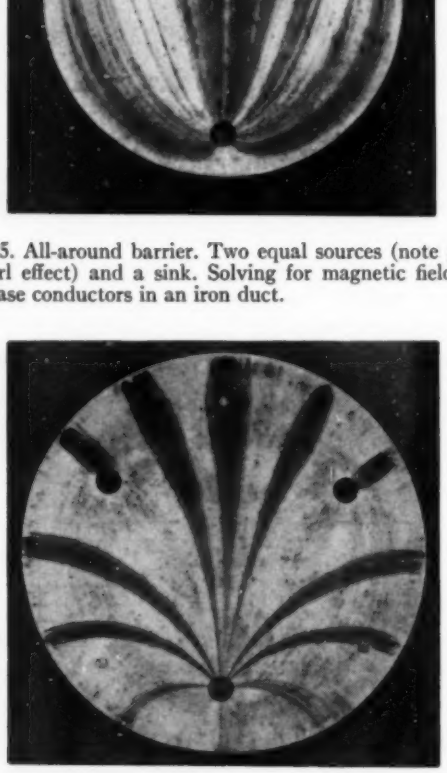


FIG. 36. Sandbed operating as sink, using shotted nickel. Crystals attached to plate in patches, showing how certain pattern effects can be emphasized.

the Foucault or damping currents would be in the system orthogonal to the fluid flow lines shown here. The theory requires a double substitution. First, the new flux-band-and-pencil-pair concept (Moore[3]) is applied; next, two narrow sectorial sandbeds replace the pencil pairs.

14. SANDBED MATERIAL

The first (1943) demonstrators were made of wood, and used Standard Ottawa sand. The fluid was air. It was loaded with ammonia fumes, to react with hydrochloric acid dotted on the lower surface of the plate and make smoke. The slab surface and the sand were blacked with India ink, to make the smoke visible. But smoke diffuses rapidly, and it may soon clog a sandbed.

The writer has successfully used sand under water, but cannot recommend it. It is too prone to float, too hard to see, and too easily washed out by placing the plate.

Shotted metal is far superior to sand. Shotted copper —also called copper shot, copper powder, copper granules, is spheroidal. All but two of these sandbeds are of copper shot, about 0.02 in. in diameter, kindly supplied as a sample by Metals Disintegrating Company, Inc. (Copper Powder MD34A). This company recently furnished a sample of shotted nickel, slightly smaller than the copper and more spherical: see the lower hole in Fig. 36. (However, the nickel is experimental, and is being closed out.) The beds of Figs. 31 and 32 were larger than the copper supply; as a stopgap, shotted cast iron, copper-plated, was used. Copper darkens rapidly with use, but is brightened again in an acid bath. Nickel is darker, but stable.

15. THE SCREEN

The screen, placed $\frac{1}{2}$ in. down, is a two-piece affair. The round holes seen in Fig. 10 are in the lower member, made of No. 22 gauge stainless steel perforated plate. The upper member is 70-mesh brass or bronze screen. They are in contact. Handwork must be done on both, after cutting, to flatten them. The plate gives flatness and strength; the screen holds the shot. Rigid mounting is required. In Fig. 2, note the enlargement of the metal core, to provide a $\frac{1}{8}$-in. enlargement of the bottom of the source chamber. The screen fits against the offset. It can be plastered at the edges to hold it; or a metal frame can bear against it, with a plate soldered to the frame to seal the bottom. The frame is plastered in.

16. MAKING THE SANDBED

The sandbed must be built under water. When water is brought into a dry-made bed, further settling occurs, and some of the surface will float away. Entrapped air is Enemy Number One. Air will always be trapped under the screen, after immersing the slab. Every bit of it must be sucked through, using the end of a rubber

[3] A. D. Moore, "Eddy currents in disks: driving and damping forces and torques," Trans. A.I.E.E. **66**, 1–11 (1947).

tube as a siphon, and rubbing it all over the screen. The wetted shot is kept at hand in a beaker. Plenty of detergent, such as Aerosol or Tergitol, is used in the water in the beaker, and detergent must be used in the tray water. Even then, shotted metal will sometimes float and cause trouble; in that case, add more detergent. After spooning enough shot into the bed to fill it, roughly, a probe is inserted all over the bed to water-settle it. If this is not done, packing effects, especially at walls and corners, will leave local voids thereat, causing an unwanted extra flow at the sandbed edges. Proper settling cannot be obtained with angular shot, as was found by trying shotted zinc and shotted bronze. Copper works well; nickel, being smoother and rounder, is still better.

Shot getting out on the slab is pushed back to the bed with the edge of a piece of Celluloid. Every such grain must be taken care of. As grains are added or taken away to make the surface nearly right, some tedious work is faced. The plate is placed directly onto the slab, to find high and low areas. When just right, moving the plate will slightly move most of the top grains; also, localized grain arrangements, with grains in circular arcs, will form and be recognized.

A single loose grain on the slab will hold up the plate, resulting in a high sandbed. Figure 28 shows a high sandbed, which was due to another cause: this slab had been eroded in the central area. When enough grains were added to touch the plate, the sandbed surface was above the slab. The streamlines are refracted, tending to avoid the bed. Such a test should be made of any important sandbed. In Fig. 29, the flow space has been doubled: less refraction occurs, showing that the larger the spacing, the less the effect due to spacing inequalities.

In a given flow space, the resistance varies inversely with the cube of the spacing. In a given sandbed, it varies inversely with the square of grain diameter (assuming like kinds and relative sizes of grains). Therefore, there is plenty of design latitude for building sandbed mappers, in seeing to it that sandbed resistance is always high compared to flow space resistance.

The idea of fixing a sandbed, to avoid the nuisance of building it anew each time, is seemingly attractive. However, the writer has gone to considerable lengths to develop cemented types of fixed beds. Due to packing and other effects, no success was achieved. Permeable materials such as alundum, fritted glass, etc., are available; but there are problems of cutting to size, cementing in, and plugging of pores in operation. No generally good answer to the fixed-bed problem is anticipated.

17. ISOLATED SANDBEDS

A rectangular, isolated sandbed is shown in Figs. 31 and 32. Since no slab is present, ears are attached to the metal frame, with pads. The spacers are laid on the pads to support the plate. The top of the frame is beveled all around to give a sharp edge.

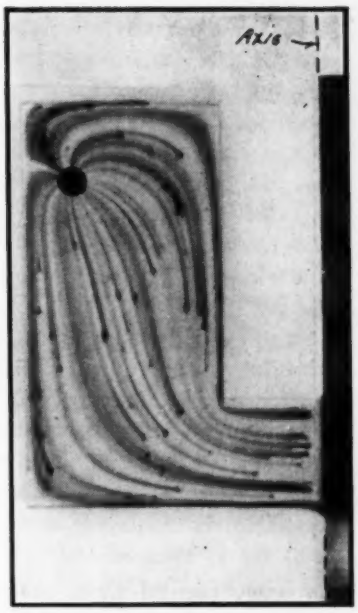


FIG. 37. Three-dimensional, with axis of symmetry. Flow space cut into flat slab by end-milling and hand scraping. Note curved mouth, to take care of edge effect.

Figure 31 shows the bed as a source, with crystals placed by sprinkling directly on the bed. Much sharper lines can be obtained, Fig. 32, by using the bed as a sink, and applying the crystals differently. Before the plate is placed, its lower surface, when dry, is painted with cellulose acetate (clear "nail polish" at any drugstore). This is applied in a band, to locate at the sandbed edge. Crystals are sprinkled on it while it is sticky. The streamlines thus formed are high in the flow space. They carry far, and are sharp.

These flow lines duplicate heat flow lines in an axially long solid having such a section, when heat is uniformly produced throughout the volume, and the surface is an isothermal. Or they simulate the equipotentials for the eddy currents in a rectangular disk uniformly and

FIG. 38. Three-dimensional of the other type, with axis of symmetry. Note gauze to kill swirl effect.

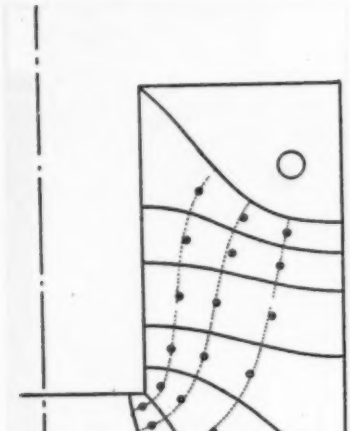


FIG. 39. Mapping study of the case of Fig. 37, to check its accuracy. If strictly accurate, the points shown would fall on their respective dotted lines.

completely pierced by a changing flux; or for eddy currents in a lamination of like section, granted constant permeability and sufficiently low frequency.

The orthogonal system would normally be secured by graphical work. However, the writer has proved, theoretically, that this system occurs as soap film contour lines. To demonstrate, a soap film was placed on the top of the empty frame, and blown up to a maximum elevation of $\frac{1}{4}$ in. at center, the frame being 2 by 8 in. The points shown in Fig. 33 were obtained (in one quarter only) with apparatus devised by the writer. Note that the four contour lines (the boundary being one) are reasonably orthogonal to the flow lines. Soap films have seen some use in solving for stresses in shafts of odd sections. Since soap film techniques are notoriously difficult, it may be that in some cases they can be discarded in favor of sandbed techniques.

18. EDGE EFFECT

In a number of the figures, the flow space is constrained, or even entirely contained by barriers placed on the slab. In Figs. 18–21, the flow is entirely confined. These lines simulate the equipotentials in a field set up by currents in conductors, entirely surrounded by iron of infinite permeability. Or, think of the fluid flow as simulating current flow in sheets shaped like the confined spaces shown, flowing between electrodes attached to the sheets. In this last case, as we approach the edge of a sheet, current density retains a finite value. But in the fluid flow space, as we approach the edge or barrier, velocity falls to zero. Here is a lack of similitude, and it must be discussed.

Interestingly, we can call on and use the sandbed mapper of Figs. 31 and 32 to help the analysis. Taking up Fig. 33 again, the writer has been able to prove that, along with other interpretations, the isopressure lines of the fluid case (or the soap film lines) coincide with isovelocity lines for axial streamline flow occurring in a long duct having this rectangular section. Therefore, we have here a picture of the fluid flow taking place near the edge of a flow space. Note that very largely, the effect of the edge is confined to a square that can be fitted into the flow space at the edge. Using data from Fig. 33, the writer has found the average velocity for the duct; and has found that the same total flow would occur, were there no edge effect, in a rectangular duct made narrower at each side by an amount equal to 0.27 times the spacing. Adopting $\frac{1}{4}$ in place of 0.27, this means that approximately, when cutting barriers, if the flow space at any edge is made *larger* than the boundaries being imitated, by $\frac{1}{4}$ of the spacing, edge effect on the fluid pattern within the simulated boundary should be very small. In a great many cases, edge effect can be ignored.

19. STUDY OF A SANDBED CASE

Interpret Fig. 25 as heat flow, heat production being distributed uniformly throughout the sandbed part of the solid. Below, a perfect heat barrier blocks all heat flow. Otherwise, all heat produced flows to the isothermal boundary (ABCD, Fig. 27), which is simulated by the open slab edge. This slab was badly eroded from previous use. Presence of metal inserts made it very difficult to grind its face flat again, and its flow space lacked uniformity by several mils. Another defect came

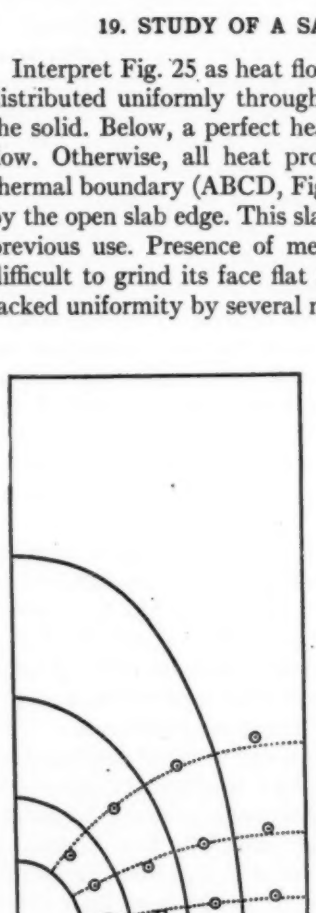


FIG. 40. Mapping study of case of Fig. 38. If accurate, the points shown would fall on their respective dotted lines.

in when the film was projected to get an enlarged full-size image for hand-copying the flow lines: the film warped due to heat. Completing the map of Fig. 27 by graphical techniques would introduce other small imperfections. Therefore, we can by no means expect the map to check perfectly.

Tracing paper was put down over the hand-copied picture. Starting at E, isothermal EF was put in, orthogonal to the flow lines. Arbitrary flow lines were next adopted and drawn in, thus dividing the total heat flow from EF to ABCD into 18 arbitrary, unequal, tubes of flow. The curvilinear squares were then constructed. By finding the conductances of these tubes, adding them, and dividing by 8, the total flow was divided into 8 equal parts, thus locating the flow lines that are drawn to the kernel. These lines divide the heat source or the sandbed into 8 areas which, in a perfect solution, would be equal. Actually, the areas, as read directly from the planimeter (starting at lower left and going clockwise) are: 0.75, 0.80, 0.70, 0.73, 0.70, 0.61, 0.62, 0.64. The largest departure from average is 15 percent. The errors largely arise from flow space inequalities. With a new slab, uneroded, far better accuracy should result.

Analysis was carried further by putting in the two isothermals GK and LP in the right half, which cross the sandbed or heat source. In the "isothermal tube" thus laid out, the dotted midline was put in; this being the isothermal that is halfway in temperature between the two other isothermal temperatures. The parts of the four heat source areas inside this midline can be taken, approximately, as dictating the heat flow that must be conducted out from GH to LM, HI to MN, and so on through four conductances. The four conductances are found by mapping the areas with curvilinear squares (not shown). They are, going clockwise, 2.00, 1.36, 1.03, 1.45. The source part-areas inside the midline, same order, are 0.70, 0.59, 0.38, 0.58. Respective ratios of conductances to source areas should all be the same. Actually, they are: 2.86, 2.31, 2.72, 2.50. The theory is approximate, and other imperfections have been mentioned above.

Even with its errors, this map is a fine first approximation to the solution, as compared with a first approximation obtainable by methods available heretofore; and with due care taken with techniques, it can be made far better.

20. THREE-DIMENSIONAL TECHNIQUES

The writer has developed theory (not included here) which shows that a fluid mapper should be able to simulate a three-dimensional field having an axis of symmetry, providing the flow space varies either with the one-third power of the radius, or inversely as the same power—depending on which of the orthogonal systems is to be shown by the fluid flow. Both types have been tested.

In Fig. 37, of the "cross-axis" type, the axis is at the right. This could be an ironclad magnet with a central pole, and circular turn carrying the current. Flow space varies inversely with the one-third power of the radius, rising to infinity at the axis. This flow space was made by starting with a flat slab, then end-milling, then hand scraping. Near the axis, spacings become large, and edge effect is large: note the curves on the plaster mouth, cut to eliminate edge effect as outlined above. A mapping study was made, Fig. 39. Each of the tubes (between solid lines) was made to predict the positions of the equipotentials that would divide this total mmf into four equal parts. In a perfect solution, the encircled points would all fall on their respective dotted lines. They very nearly do so.

The other, the "pro-axis" type, is represented by Fig. 38, where the axis is at the right. Here, the flow space varies directly with the one-third power of the radius. The mapping study, Fig. 40, again shows a very good check. This case simulates heat flow in a solid cylinder, when most of the upper half of the round surface is an isothermal, most of the lower half is another isothermal, and the middle band between takes temperatures as dictated by the conditions. No heat flow is crossing the cylinder ends. It is clear that such types of mappers will also apply to electrostatic and magnetic fields.

As far as is known, these are the first such fluid mappers ever to be produced.

21. REFRACTION

When the slab in Fig. 22 was cast, a square of paraffined bristol board was first placed on the plate glass. It is seen, still imbedded, in this picture. In Fig. 23, the insert has been removed, leaving a depression, and demonstrating refraction when a sharp change of flow space occurs. The lack of symmetry shows that this technique for making a depression has its imperfections. However, the point is that, as was pointed out long since by Hele-Shaw,[4] change of flow space, properly effected, can be made to correspond to change of permeability; it also corresponds to change of thermal conductivity, dielectric constant, and so on.

In Fig. 12, for example, the sandbed was level with the slab, and no refraction occurs at the sandbed boundary—merely a change in curvature; and here, the properties of the simulated material are the same, inside the source and outside it. But by inserting a thin lining around this sandbed and elevating it, a high sandbed could be built. Then, in terms of heat conduction, the source volume would be given a lower thermal conductivity than the surrounding solid. Like-

[4] H. S. Hele-Shaw and Alfred Hay, Phil. Trans. Series A (English) **195**, pp. 303–328 (1901).

wise, it is possible to build depressed sandbeds. Other variations are possible, given the incentive.

22. TEACHING POSSIBILITIES

A great many of the simpler fluid mappers can be made up easily and cheaply. Their visual appeal is very high indeed. There is no reason why these teaching aids should not be adopted wherever a better understanding of field situations can help students, at any level.

Note added in proof.—Since writing the paper, it has been found: (a) that a line of Silicone DC4 applied to the dry lower side of a rubber barrier and gently pressed by the plate, makes a good seal; and (b) excellent 35 mm color slides (Ansco Daylight) are made, using blue photofloods.

High Inverse Voltage Germanium Rectifiers*

S. BENZER**
Department of Physics, Purdue University, Lafayette, Indiana
(Received January 17, 1949)

Rectifying current-voltage characteristics going up to several hundred volts inverse have been observed in metal-germanium point contact rectifiers. A reproducible negative differential resistance region occurs in the inverse characteristic. Certain impurities are desirable in producing high voltage material. Surface treatment, e.g., by etching, is very important. The metal used as a whisker has little effect. Increasing the force of contact increases greatly the current of low voltages but has less effect on the high voltage curve. Pronounced improvement of rectification can be effected by treatment of the contact with large currents.

Variation with temperature is very marked, especially for crystals of large inverse resistance; the variation of the inverse peak with temperature indicates that contact heating is responsible for the negative resistance. Time lags in the inverse negative resistance region of the order of 10^{-5} second occur.

When contact is made between two Ge crystals, typical inverse characteristics are observed in both directions. Photoelectric effects are observed and indicate that the barrier thickness is greater the higher the inverse peak voltage.

I. INTRODUCTION

CONTACT rectification between germanium crystals and metals has been known since the work of Benedicks.[1] Merritt[2] also did extensive work on this problem. In 1942, when germanium rectifiers were found, by the Sperry Research Laboratories, to be useful at microwave frequencies, a group at Purdue, working under the direction of K. Lark-Horovitz (contract with NDRC Division 14) undertook to develop such crystals for use in radar.

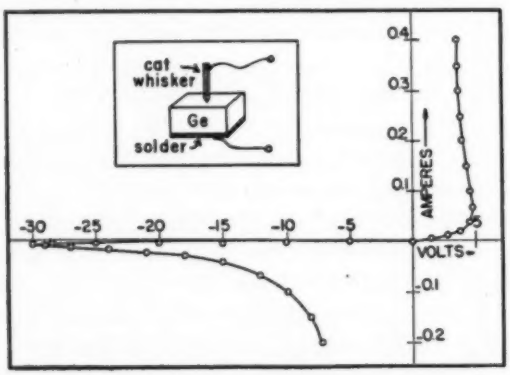


FIG. 1. Double-valued current-voltage characteristic for metal-germanium contact (inset). Polarity of the applied voltage refers to the metal.

In the course of the investigation of these rectifiers, it was necessary to determine the characteristics of germanium in "burn-out," as the destruction of rectifying properties by excessive power is called. When more than a few volts were applied to crystal rectifiers made with germanium, interesting current-voltage curves were observed (1942) which became the subject of this study. Figure 1 is an illustration of the type of characteristic which is referred to. As the current through the crystal was increased, the voltage would go through a maximum and then drop again, thereby giving a negative resistance characteristic which persisted as negative up to high currents. This effect was most marked in the high resistance or inverse direction (metal negative with respect to the germanium), but also could be observed in the forward, although at voltages and currents widely different in the two directions.

There were uses for crystals which would maintain a high resistance up to large voltages of one polarity, while having a low resistance for the other polarity. Thus, the investigation of the properties of these crystals and the mechanism responsible for the high inverse voltage characteristic became of interest. During studies on the effect of air on contact rectification in germanium, Lark-Horovitz and Whaley[3] found that

* Based in part on Purdue University NDRC report 14-323, November 1, 1944, entitled, "The high voltage germanium rectifier" and on a thesis submitted in partial fulfilment of the requirements for the M.S. degree at Purdue University. This is paper No. 8 of a series of studies from the Purdue Semiconductor Laboratory.

** At present on leave of absence at Oak Ridge National Laboratory.

[1] C. Benedicks, Int. Zeits. f. Metal. **7**, 225 (1915).

[2] E. Merritt, Proc. Nat. Acad. Sci. **11**, 743 (1925).

[3] See final report (reference 4), also R. M. Whaley and K. Lark-Horovitz, Phys. Rev. **69**, 683 (1946); and R. M. Whaley, Ph.D. Thesis, Purdue University, (February, 1947).

some germanium melts withstood very high inverse voltages. This led to a systematic investigation of germanium alloys which would produce good high voltage rectifiers. The results of this project are summarized in the final report by K. Lark-Horovitz.[4]

Since semiconductors generally have negative temperature coefficients of resistivity within certain temperature ranges, a zero or negative dynamic resistance may occur due to heating. Effects of this type were observed in coherers, which may have involved thin layers of semiconducting oxide.[5] Lyle,[6] using a slab of boron, and Waterman,[7] using strips of molybdenite, observed negative resistance and breakdown phenomena. Others[8] found negative resistance in contacts between metals and various crystals. Thermistors,[9] which consist of sintered mixtures of semiconducting oxides, exhibit pronounced negative characteristics.

All these formerly observed characteristics were similar for both polarities of applied voltage, and if rectification effects existed, they were slight. In germanium crystals, however, the great difference for the two polarities, and the ability to withstand large voltages in the inverse direction are distinguishing features. By addition of certain impurities, similar characteristics have been obtained with silicon.[10]

II. EXPERIMENTS

A. Apparatus

The apparatus involved in these experiments was in general very simple, consisting merely of mechanical arrangements for holding a metal wire (cat whisker) in contact with a germanium crystal and of instruments for measuring electrical characteristics. When it was desired to probe around over the surface of a crystal to observe characteristics at different points, the germanium crystal was soldered directly (using Nokorode flux) to a brass stud which fit snugly into the platform of a test stand. The crystal could be moved in two directions, and the force on the whisker could be varied by a calibrated spiral spring. A standard cartridge mount*** was used when a small fixed unit was desired. In this arrangement, a wafer of germanium about 1-mm thick was soldered to a brass rod on one end of the cartridge. A crimped cat whisker made contact from the other end. Some cartridges were filled with wax for waterproofing and reasons of mechanical stability, and some were fitted with pigtails for convenience in wiring.

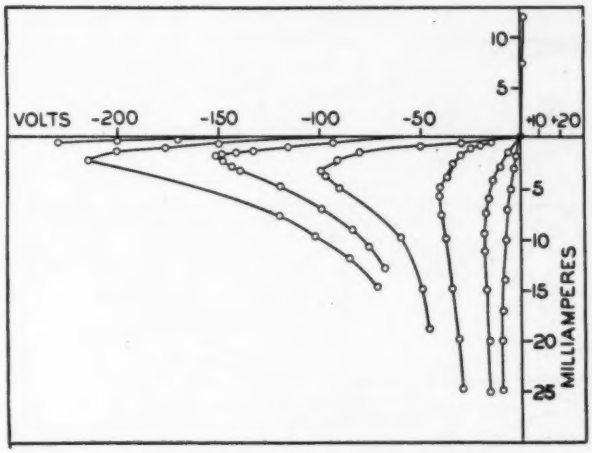


FIG. 2. Inverse high voltage d.c. characteristics.

For measurements of high voltage characteristics at various temperatures, a thermocouple was imbedded in the solder to indicate the base temperature of the germanium crystal, so that compensation could be made for the power dissipated at the contact.

Whiskers, unless otherwise stated, consisted of 0.005-in. diameter tungsten wire, sharpened electrolytically using alternating current and a 25 percent KOH solution. The conical points were uniformly sharp with tip diameter less than 0.001 in.

For exploring the surface of a crystal with a whisker, a simple circuit was used, which instantaneously plotted the current voltage curve on an oscilloscope. In the case of negative resistance characteristics, the plate resistance of a thermionic tube was used to preserve stability, the current through the crystal being varied by grid modulation. By using a small a.c. swing and raising the d.c. plate current until the trace on the oscilloscope was vertical, the "peak" voltage (i.e., the voltage for zero differential resistance) could readily be determined.

B. Experimental Results

1. D.C. Characteristics

(*a*) *High voltage.*—The various parameters of the d.c. characteristic such as forward resistance, back resistance, and peak voltage vary greatly from one germanium crystal to another, depending markedly on factors such as the composition, the surface, and treatment of the rectifying unit after assembly. Figure 2 shows a series of d.c. characteristics for various crystals, illustrating the wide variation obtainable in the peak inverse voltage. The polarity of rectification in these crystals is such that a larger current flows when the germanium is negative (*N*-type rectification). Note that the value of this peak voltage varies from a few volts to several

[4] K. Lark-Horovitz, "Preparation of Semi-Conductors and Development of Crystal Rectifiers," Final Report, NDRC 14–585, Purdue University (covering March, 1942 to November, 1945).

[5] A. H. Taylor, Phys. Rev. **16**, 199 (1903); A Fisch, J. de Phys. **3**, 350 (1904); W. H. Eccles, Proc. Phys. Soc. (London) **22**, 289, 360 (1910).

[6] F. W. Lyle, Phys. Rev. **11**, 253 (1918).

[7] A. T. Waterman, Phil. Mag. **33**, 225 (1917); Phys. Rev. **21**, 541 (1923).

[8] O. W. Lossew, W. World (October 22, 1924) p. 93; Zeits. f. Fernmeldetechnik **6**, 132 (1925); F. Seidl, Zeits. f. Phys. **27**, 64 (1926) and **27**, 816 (1926); K. Lichtenecker, Zeits. f. tech. Physik **8**, 161 (1927); K. Sixtus, Zeits. f. tech. Physik **9**, 70 (1928); F. W. Kallmeyer, Ann. d. Physik **86**, 547 (1928); E. Habann, Ann. d. Physik **9**, 1 (1931).

[9] Becker, Green, and Pearson, Trans. A.I.E.E. **65**, 711 (1946).

[10] Lewis, Taylor, Gibson, and Stephens, "High Back Voltage Silicon," NDRC Report 14–453, University of Pennsylvania (June 28, 1945).

*** See reference 14, p. 16.

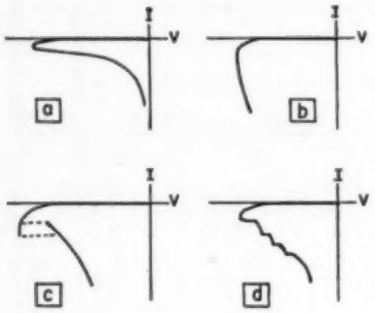


FIG. 3. Varieties of inverse characteristic which may be observed with an oscilloscope.

hundred.**** It is remarkable that the contact between two conducting materials will support such large voltages with small current flow for one polarity, while current flows freely when the voltage is reversed.

Such a curve is quite reproducible provided it has previously been traversed at least once. As the current is raised for the first time, progressive changes occur in the contact. However, subsequent measurements up to currents no larger than that previously applied, give reproducible results.

The curves in Fig. 2 form a family, the current at the peak being larger for small peak voltages. These curves have essentially the same shape, as may be seen by replotting them on different scales to make the peaks coincide. The product of voltage and current at the peak is of the order of 0.3 w, which is evidently sufficient to cause a considerable rise in temperature of the crystal mount. Therefore, in order to obtain a curve which repeats for rising and falling current, one must either perform the measurement slowly enough to allow for thermal equilibrium or rapidly enough so that no appreciable heating of the crystal mount takes place.

Careful measurements at constant temperature, cooling and heating the crystal while a thermocouple attached to the bottom of the crystal was maintained at constant reading, showed that the negative resistance effect in the back direction still persists.

Not always is the high voltage characteristic as simple as those in Fig. 2. As shown in Fig. 3 *a* and *b*, the degree of the negative effect may vary considerably from one crystal to another. Curve *c* illustrates a common case, where in spite of the high impedance circuit, discontinuous jumps are observed. This type of characteristic is double-valued in both voltage and current. Curve *d* illustrates a sort of stepwise breakdown which sometimes occurs. Curve *e* shows a case with a double peak. As illustrated by curve *f*, negative resistance oscillations may occur spontaneously while the characteristic is being observed.

The high voltage characteristic in the forward direction also varies from one sample to another, and correlates roughly with the bulk resistivity of the germanium. The peak voltage may vary from 2 v for a low resistivity to 25 v for relatively high resistivity germanium. In some cases, the peak with the whisker positive may occur at higher voltage than the reverse.[11]

Due to the very large currents involved, one must be especially careful in measuring forward characteristics to allow for the heat developed. When care is taken to keep the base temperature of the crystal constant, the negative resistance is in most cases greatly reduced.

(*b*) *Low voltage characteristic.*—Detailed experiments by Yearian[12] and Smith[13] on characteristics of low inverse voltage germanium rectifiers have shown that the metal-germanium contact may be approximately represented by a model consisting of a voltage dependent resistance R_c due to a barrier layer formed at the contact, a capacitance C (voltage dependent) which shunts it, and a spreading resistance R_s due to the germanium in series.

According to the simple diode theory of rectification,[14] R_c is approximately given by a relation of the form $i=i_0(\exp\alpha V-1)$, where i_0 and alpha are constant at a given temperature. For a perfectly uniform contact, the value of alpha should be e/kt or 40 v^{-1} at room tempera-

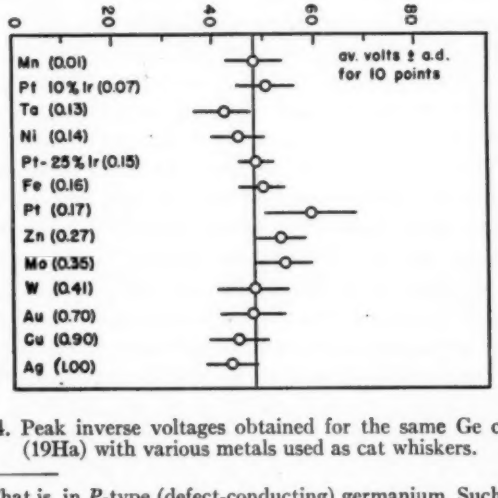


FIG. 4. Peak inverse voltages obtained for the same Ge crystal (19Ha) with various metals used as cat whiskers.

**** Theurer and Scaff (reference 22) have reported over 400 volts!

[11] That is, in *P*-type (defect-conducting) germanium. Such crystals are usually non-rectifying at low voltage, but are unsymmetrical so far as their peak voltages are concerned. R. Bray and K. Lark-Horovitz, Phys. Rev. 71, 141 (1947) found that with proper etching and choice of cat whisker, rectifying and photosensitive *P*-type crystals could be made. See also W. C. Dunlap and E. F. Hennelly, Phys. Rev. 74, 976 (1948).

[12] H. J. Yearian, "Investigation of Crystal Rectifier DC Characteristics," NDRC Report 14-115, Purdue University (December 3, 1942).

[13] R. N. Smith, "Crystal Capacity as a Function of Bias and Its Relation to the Theory of Crystal Rectification," NDRC Report, Purdue University (March 16, 1946).

[14] For a detailed review of the theory of rectification, see H. C. Torrey and C. A. Whitmer, *Crystal Rectifiers* (McGraw-Hill Book Company, Inc., New York, 1948).

ture. Johnson, Smith, and Yearian[15] have shown that fluctuations in barrier height over the area of contact lead to smaller values of alpha. According to potential theory, the value of R_s should be $(\rho/4a)$, where ρ is the resistivity of the germanium, and a is the radius of contact, provided the resistivity is independent of electric field and temperature. The capacitance C is given as $C=\epsilon A/4\pi d$, where ϵ is the dielectric constant, A is the area of contact, and d, the thickness of the barrier, is equal to $[\epsilon(\phi-V)/2\pi Ne^2]^{\frac{1}{2}}$, ϕ being the height of the barrier, and V the applied voltage.

The usual way of analyzing low voltage characteristics[12] is to plot log$i$ *vs.* applied voltage. The lower part of the curve is adjusted by using trial values of i_0 until a straight line is obtained. Similarly, trial values of R_s are used to correct the upper part of the curve by subtracting iR_s from the applied voltage.

In general, this analysis is not successful when applied to high inverse voltage germanium rectifiers due to the complex nature of the spreading resistance. The latter does not seem to be independent of voltage or frequency. Bray, Lark-Horovitz, and Smith[16] have shown that the spreading resistance may be a factor of twenty or more lower than the value computed from data on bulk resistivity. Also, Yearian[17] has shown that the spreading resistance increases with frequency. Actual data on low voltage d.c. characteristics are given below in various sections.

2. *Preparation of Crystals*

(*a*) *Composition of germanium alloy.*—Germanium, a semiconductor, has relatively high resistivity (of the order of 60 ohm-cm at room temperature†) in the pure state. In order to make it suitable for use in rectifiers, impurities must be introduced, which create disturbances in the lattice yielding free electrons (or holes, depending upon the impurity) and raising the conductivity.

The samples of germanium with various added impurities were prepared by R. M. Whaley.[18] Among the alloys which produced good rectifiers there were two main types: (1) those which had high back peak voltages but dropped greatly in rectification efficiency at microwave frequencies, (2) those which continued to rectify well at centimeter wave-lengths but would not maintain more than a few volts inverse. In general, the high voltage crystals were also quite noisy compared with low voltage ones.

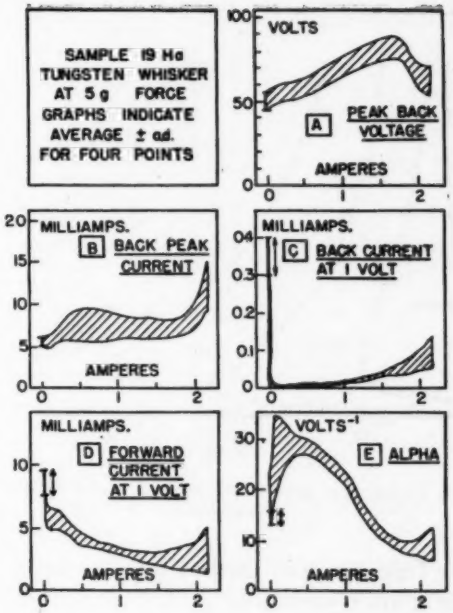


FIG. 5. Variation of characteristic parameters with progressively larger amounts of current previously passed through the contact.

If insufficient impurity is added, the conductivity will be too low. Excess impurity tends to deposit at grain boundaries, producing either shorting or insulating layers. In the case of impurities such as calcium, these layers may oxidize with time, causing the crystal to crumble. The fact that crumbling down to very small grains occurs, illustrates the degree to which such layers may penetrate through the sample.

(*b*) *Surface treatment.*—As one might expect, the characteristics observed depend very strongly upon surface treatment. A freshly broken surface will, in general, be found to give perfectly good rectification, provided, of course, that one starts with a good germanium sample. Such broken surfaces are very smooth and the cat whisker (even tungsten) slides all over, making it difficult to assemble a permanent unit. In practice, the surface may be ground flat, using 600 mesh carborundum under water, following up with Al_2O_3 and emery paper if desired. This greatly reduces the back peak voltage. However, by subsequent etching, the rectification may be restored to even better than that for the freshly broken surface. Thus, crystal 19 Ha, broken open and contacted with a cat whisker, gave inverse peak voltages around 70 volts. Upon grinding and polishing, this value dropped to 20, but went up to 100 volts after etching for one minute in an $HF-HNO_3-Cu(NO_3)_2$ mixture.[18,19] Electrolytic etching offers some advantage over chemical in that it will show up inhomogeneities in the electrical properties of the crystal by preferential etching of low resistance

[15] Johnson, Smith, and Yearian, "Semi-Quantitative Explanation of DC Characteristics," NDRC Report, Purdue University (August 14, 1943).

[16] Bray, Lark-Horovitz, and Smith, Phys. Rev. **72**, 530 (1947).

[17] H. J. Yearian, "Dependence of Forward Conductance and Back Resistance of High Voltage Germanium on Voltage and Frequency," NDRC Report 14-581, Purdue University (October 31, 1945).

† This value is obtained from extrapolation of resistivity *vs.* temperature data (reference 24).

[18] R. M. Whaley and P. Pickar, "Preparation of High Voltage Germanium Crystals," NDRC Report 14-341, Purdue University (November 1, 1944). For further details on preparation, see references 4 and 14.

[19] Boyarsky, Pickar, McDonald, Smith, Whaley, and Yearian, "Production and Performance of Germanium High Back Voltage, High Back Resistance Crystal Rectifiers," NDRC Report 14-577, Purdue University (October 31, 1945).

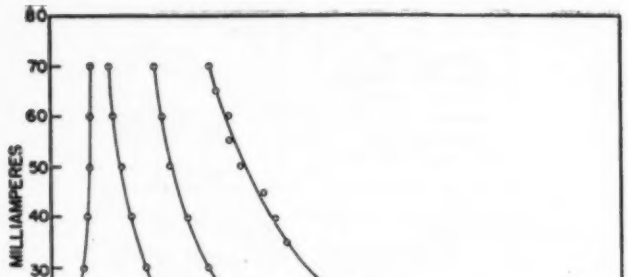


FIG. 6. Variation of inverse characteristic with temperature (crystal 9*Ha*).

regions. This is a good method for locating rectifying layers at grain boundaries.

(*c*) *Effect of whisker.*—The back direction peak voltage for a given sample depends very little upon what metal is used as whisker, at least for the thirteen metals that were tried. Each metal was tested at ten different points on a small region of a relatively uniform sample. The load was adjusted so that the forward current at one volt was 20 ma, thus assuring that the contact areas were approximately the same. Figure 4 shows the average values and also the average deviations observed for each metal. The thirteen metals are arranged in order of increasing thermal conductivity and there seem to be no differences from one whisker to another which are greater than the fluctuations with one type of whisker alone. When the whiskers are arranged in order of increasing work function there is also no trend evident. Choice of a whisker material should thus be made on the basis of requirements other than the peak voltage desired.

As is seen in a later section, one obtains about the same peak voltage even if another germanium crystal is used as the whisker.

(*d*) *Effect of contact force.*—The effective area of contact may be varied by changing the force pressing the whisker against the germanium. A tungsten point flattens out to an area approximately proportional to the force[20] and corresponding to the tensile strength of tungsten. At a force of 10 g, a tungsten whisker gives a contact diameter of the order of 10^{-3} cm.

This flattening is irreversible and may be observed in a microscope. However, the actual contact area cannot be assumed to be equal to the apparent value. In fact, the resistance of the contact is, with some hysteresis, reversible with the force, in spite of the irreversibility of the flattening. Surface irregularities are probably responsible for the discrepancy; the intimacy of contact and hence the effective area change with the force.

Increasing the force has a very pronounced effect on the current at low voltages, whereas at high voltages the current increases much less rapidly. So far as the peak voltages are concerned, the force need not be critically adjusted. If it is desired to keep the inverse resistance high, however, too much force must not be applied. The rectification ratio is best for small force, since the back resistance goes inversely as the area of the contact, while the spreading resistance goes inversely as the radius.

(*e*) *Power treatment.*—In experiments on burnout of germanium rectifiers, progressively larger amounts of power were passed through contacts to determine the point beyond which the rectification would de-

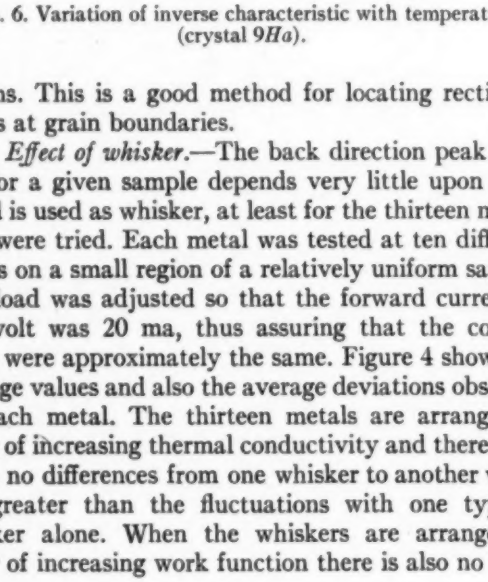


FIG. 7a. Inverse peak voltage and current at the peak *vs.* temperature.

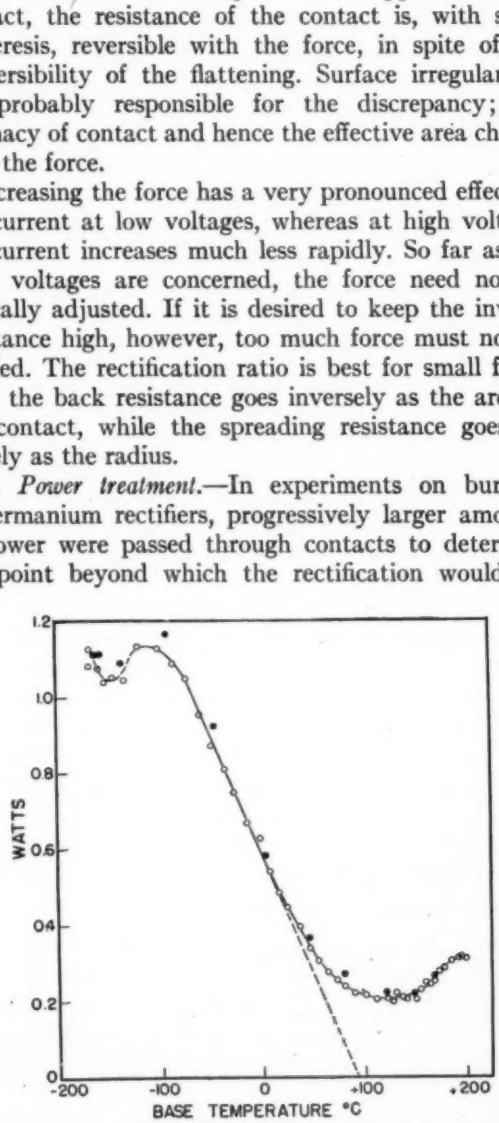


FIG. 7b. Product of peak voltage and current *vs.* temperature. (crystal 12 *N*)

[20] J. A. Becker and J. N. Shive, Report MM-42-110-24 Bell Labs. (November 11, 1942).

teriorate. It was found that the initial effect of power was a marked improvement in the back resistance of the rectifier, while further power would ruin it.[21] Essentially the same effect is produced regardless of the polarity of the current sent through the crystal, which would indicate the effect is primarily one of heating at the contact rather than an electrolytic one as is employed in "electroforming" selenium rectifiers.

The power treatment effect takes place in a fraction of a second and either a.c. or d.c. may be used. A current pulse of the order of several tenths of an ampere will generally give optimum results. In addition to improving the back resistance and peak back voltage, the power treatment greatly increases the mechanical stability of the unit. In fact, under proper conditions the whisker may even be welded securely to the germanium, making a very rugged unit.

Figure 5 illustrates the changes in the various parameters of the characteristic after progressively larger amounts of direct current are passed. Each graph represents the average of four runs with tungsten cat whiskers at different points on the same crystal, the current being passed in the forward direction for two points and in the back direction for the other two. (At the large currents involved, one is well over the peak in the back direction, so that the voltage across the crystal is comparable in the two cases.) The back current at one volt is seen to drop sharply at first, but if too much current is passed, it starts to rise again. The peak voltage increases at first. The forward current at one volt drops considerably, which is one drawback to this process.

In Fig. 5, the values of alpha before power treatment are seen to range about 15, but become quite high in the initial stages and later drop to below 10. The great increase in alpha in the early stages may be understood in terms of the multicontact theory as due to removal of spots of low contact potential. (Such spots will carry the major portion of the current.)

Note the important fact that the changes in peak back voltage do not correspond with those of alpha- and back current at low voltage. While the latter two change radically at first, the peak voltage does not follow and, in fact, continues to increase after their optimum is passed.

Certain hysteresis effects were noted at Purdue[19] and at Bell Laboratories[22] in testing high and low voltage characteristics of contacts which had not been power treated. If one first measured the back resistance at 5 v, say, and then raised the voltage to determine the peak value, a repeat measurement at five volts would give a resistance considerably lower than the first value.

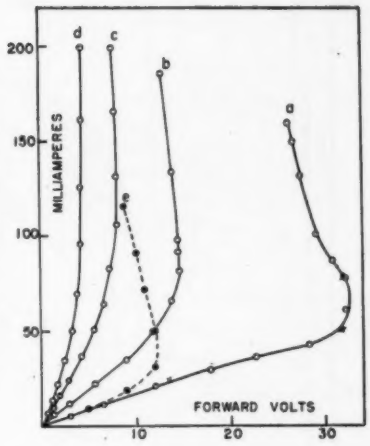


FIG. 8. a, b, c, d: Variation of forward high voltage characteristic with base temperature, keeping the base temperature constant for each curve. Dotted curve *e* illustrates result when base temperature at zero voltage is the same as in *a* but no compensation is made as the current is raised.

However, if a few milliamperes were passed in the forward direction, which required only a volt, the inverse resistance at 5 v was found to recover to its original value. Power treatment has the effect of eliminating this hysteresis to a great extent.

3. *Effect of Temperature*

(*a*) *High voltage characteristic.*—The high voltage forward and back direction characteristics are strongly temperature dependent, just as the low voltage characteristics are. In every case tested, the peak voltage (both forward and back) decreased with increasing base temperature.

Figure 6 illustrates the variation of the back high voltage characteristic with temperature. This temperature variation is perfectly reproducible, provided, as always, that sufficient power has previously been applied to the point. Mechanical conditions must also be stable, or the whisker will shift to a new point. Previous treatment with power helps in this regard also.

It is more convenient to simply measure the peak

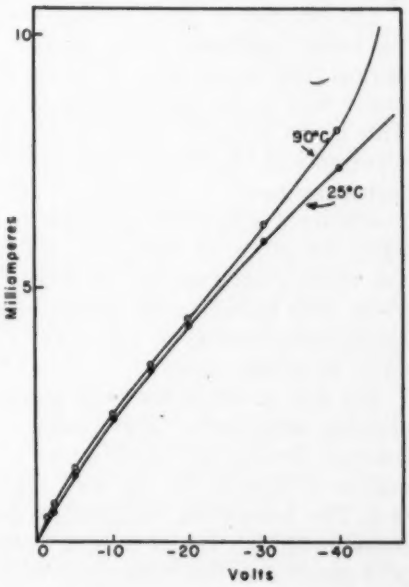


FIG. 9. Effect of temperature on the inverse characteristic of a low resistance crystal. (27 *E*-4).

[21] Similar effects were noted by Lawson, Maurer, Miller, Schiff, and Stephens, "*DC* Burn-Out Temperature in Silicon Rectifiers," NDRC Report 14–113, University of Pennsylvania (November 1, 1942).

[22] J. H. Scaff and H. C. Theurer, "Final Report on Preparation of High Back Voltage Germanium Rectifiers," NDRC 14–555, Bell Labs. (October 24, 1945).

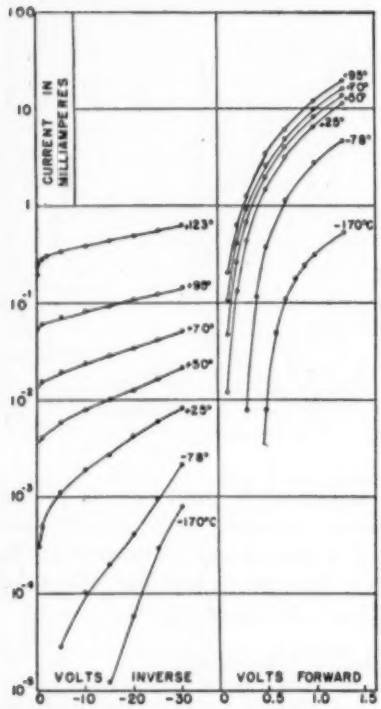


FIG. 10. Semi-log plot of inverse and forward characteristics *vs.* temperature for a high inverse voltage, high inverse resistance crystal (24 P-32).

voltage and current as a function of temperature instead of the entire characteristic. Figure 7a illustrates the variation of these parameters above room temperature. Note that the peak current goes through a minimum while the peak voltage continually drops with increasing temperature. By the time a base temperature of 200°C is reached, the back resistance is very low.

When the base temperature is reduced below zero, the peak voltage continues to increase, at least down to liquid air temperature. Open and closed circles are used to represent the two halves of the cycle with rising and falling temperature. Note the good reproducibility attainable.

The product of peak voltage and peak current, or the power dissipated at the peak, is plotted in Fig. 7b against base temperature. Note the linear drop of peak power with increasing temperature over a wide range. This is the behavior which would be expected if the occurrence of the peak corresponded to a certain contact temperature.[21] Furthermore, extrapolation to the temperature axis would give the base temperature at which no power is needed to reach breakdown, i.e., the contact temperature at which breakdown takes place. This extrapolated temperature usually is in the range corresponding to the onset of intrinsic conductivity in the germanium.

The rate at which the peak power falls off with increasing temperature corresponds, if the picture of a constant breakdown temperature is assumed, to the number of degrees rise per watt dissipated at the contact. The breakdown temperature cannot be assumed constant at high values of base temperature, since the peak power does not continue to fall.

Figure 8 illustrates the variation of the high voltage characteristic in the forward direction with temperature, for a high resistivity sample. This follows the same pattern as the back direction variation, except for the less pronounced negative resistance, which is even less evident for low resistivity crystals.

(*b*) *Low voltage characteristic.*[23]—As was quite evident in the previous sections, the characteristics at low and at high voltage are largely unrelated, since large changes may take place in one and not the other. This is true also with respect to temperature variation. So far as the temperature variation at low voltage is concerned, a crystal may be roughly classified according to the magnitude of its low voltage back resistance. For low resistance crystals, the temperature variation is small. An extreme example is shown in Fig. 9; although this crystal has a peak voltage of over 50 v at room temperature, the back resistance is only 4000 ohms. From 25° to 90°, there is little change in the low voltage resistance.

For contacts of high resistance the temperature effect is large. Figure 10 illustrates the forward and back characteristics at low voltage, plotted on a semi-log scale, for a high resistance unit at various temperatures. Note the saturation effect in the back characteristic and the very rapid change of saturation current with temperature.

This variation is similar to that expected of a simple barrier according to the diode theory. According to this theory, the back saturation current i_0 is given by $i_0 = j_0 A \exp[-(e\phi/kT)]$, where j_0 is the thermal current density in the semiconductor, A is the area of the contact, and ϕ is the barrier height. Figure 11 shows a

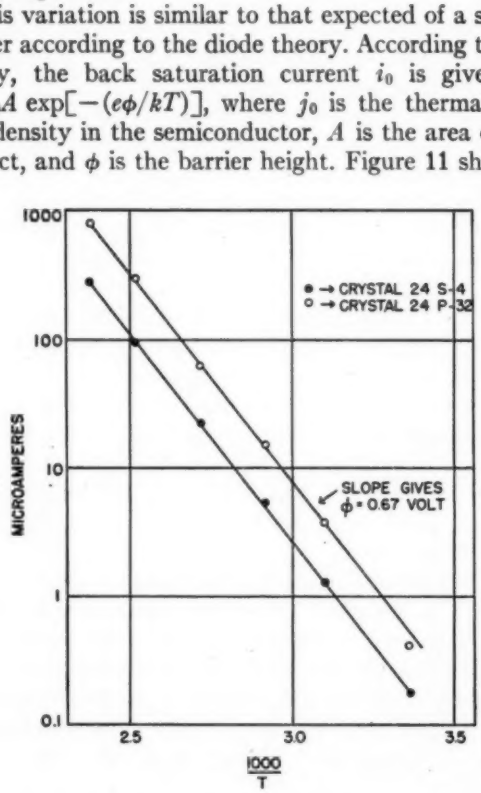


FIG. 11. Log back direction saturation current *vs.* reciprocal of absolute temperature for two high inverse resistance contacts.

[23] S. Benzer, "Temperature Dependence of High Voltage Germanium Rectifier *DC* Characteristics," NDRC Report 14-579, Purdue University (October 31, 1945).

plot of log saturation current against $1/T$, where T is the absolute temperature, for two relatively high resistance crystals. The graphs are straight lines of slope corresponding to a barrier height of over 0.6 v. This approaches the width of the forbidden band in germanium, determined from resistivity measurements[24] to be about 0.76 v. The width of the forbidden band would represent the largest barrier height attainable in germanium. While the value of ϕ is reasonable, j_0 is several orders of magnitude larger than is expected for the given semiconductor and contact area.

The forward current also increases with temperature, as may be seen in Fig. 12, the percentage variation being much larger at low voltage. The variation of alpha with temperature is interesting, since it is different from that observed by Yearian[12] for low inverse voltage units. Of the crystals he studied, Yearian found that for some the alpha was independent of temperature while for others it fell in proportion to $1/T$ as the temperature was increased. For the high voltage units here discussed, however, alpha decreases even more rapidly than $1/T$ as T is raised.

Figure 13 illustrates this variation with temperature of alpha-values computed from the forward currents at 0.1 and 0.2 v at each temperature. While alpha is not far below the theoretical value at zero degrees, it drops to 10 or 20 percent of theoretical at 145°C. Part of this anomaly may be due to the fact that the contact resistance becomes so low at high temperatures that the non-linear spreading resistance has a large effect. At very low temperatures, the currents at low voltage become very small and alpha is difficult to measure.

Note in Fig. 12 that at liquid air temperature, the

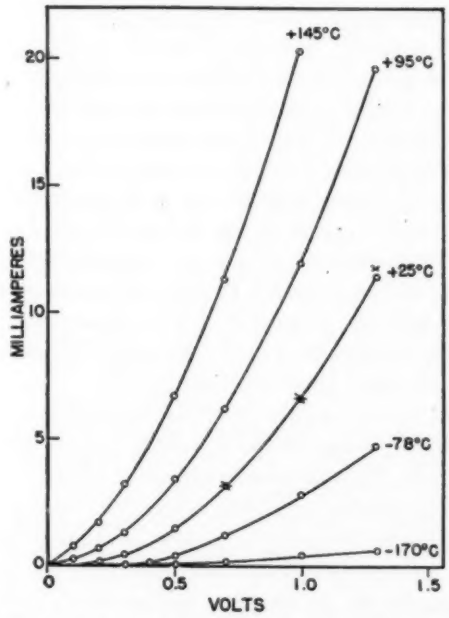


FIG. 12. Forward characteristic of crystal 24 *P*-32 *vs.* temperature.

[24] Lark-Horovitz, Middleton, Miller, and Walerstein, Phys. Rev. 69, 258 (1946).

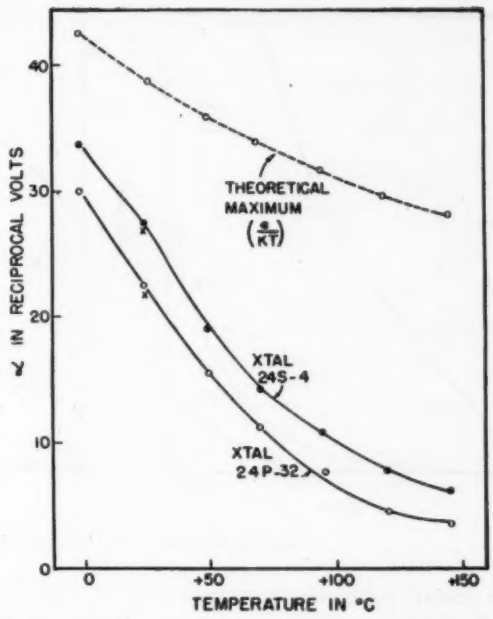


FIG. 13. Variation of alpha with temperature. Alpha is computed from the currents at 0.1 and 0.2 v in the forward direction, using the formula $\alpha = 23 \log_{10}[(i_{0.2}/i_{0.1}) - 1]$.

forward resistance is much higher than at room temperature. This cannot be simply due to an increase of bulk resistivity (hence, increasing the spreading resistance) because the bulk resistivity does not increase by any such factor in that temperature interval. The bulk resistivity, in fact, becomes less at the low temperature.

In general, when plotted on a linear scale, the inverse characteristic up to 30 v may be analyzed into three components; the total current at a given voltage is given by the sum of the three. One component is a saturation current (diode contribution) which is reached at a fraction of a volt. The saturation current tails off into an essentially linear region (ohmic contribution), the second component. The third component is one which rises rapidly with voltage. Each crystal shows these effects to different relative degrees. Moreover, the saturation and linear components vary rapidly with temperature while the third component does not. Therefore, the shape of a given crystal's characteristic varies with temperature.

Figure 14 is a linear plot of the characteristics of Fig. 10. The variation is so large that three graphs are used in order to see all the curves in detail. At very low temperature the third component is predominant. Note how little this varies from $-170°$ to $-78°$ (a factor of two in absolute temperature). At 25°, the saturation and linear components become important and at higher temperatures the third component is no longer noticeable. However, at very high temperature (148° curve in Fig. 14) the curve begins to rise once again. This is due to the fact that the peak voltage drops rapidly as the temperature exceeds 100° and, in the crystal illustrated, the peak has actually fallen to only 30 v at 148°.

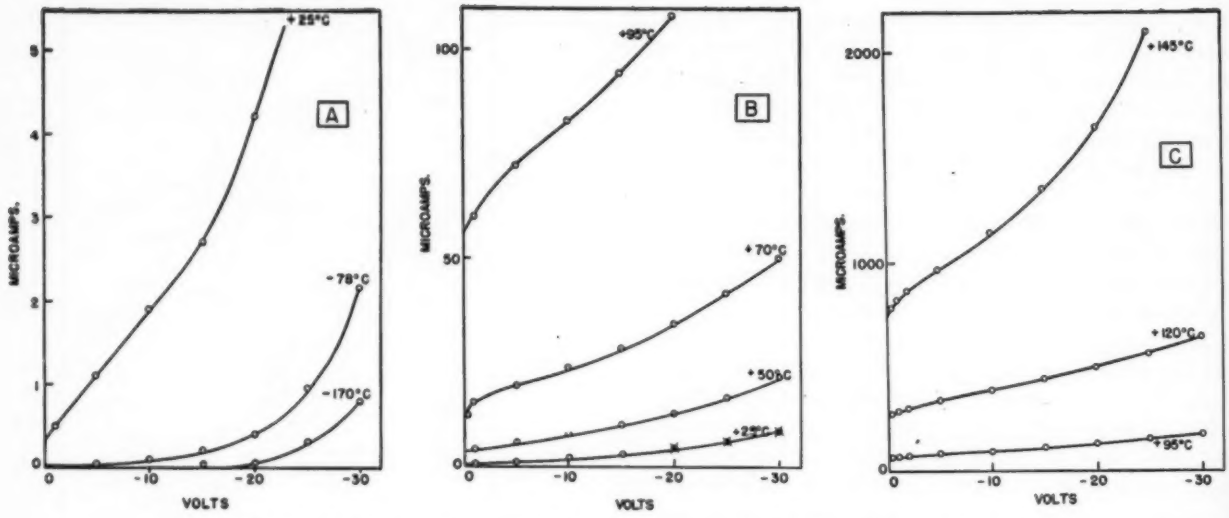


FIG. 14. Linear plot of inverse characteristics of crystal 24 *P*-32 *vs.* temperature.

The linear component corresponds, on the multi-contact theory, to the current conducted by small spots with contact potential either low or of wrong polarity. Since the diode and ohmic components vary rapidly with temperature while the peak does not, the third component is perhaps the only one which is directly related to the peak region of the characteristic.

Although factors such as the composition of the germanium alloy, surface treatment, etc., have large effects on the actual characteristic, the temperature dependence of the characteristic does not depend upon them. That is, a high back resistance crystal shows its typical temperature dependence regardless of whether the impurity is nitrogen or tin, or whether the whisker is tungsten or platinum-iridium, or whether a chemical or electrolytic etch is used. A given unit before power treatment may have a fairly low resistance and a small temperature effect. After power treatment the back resistance is increased, but the temperature effect also increases. The effective improvement, therefore, is less pronounced at high than at low temperature, but may still be quite large.

4. Time Effects

Here again it is necessary to consider the low and high voltage characteristics separately. When used at low voltage, germanium rectifiers continue to operate at wave-lengths of 10 cm and less. However, as pointed out earlier, rectifiers made of alloys which give high inverse voltages are relatively poor. The response of such crystals drops above a frequency of about 100 Mc and is very small, though still detectable, at 3000 Mc.

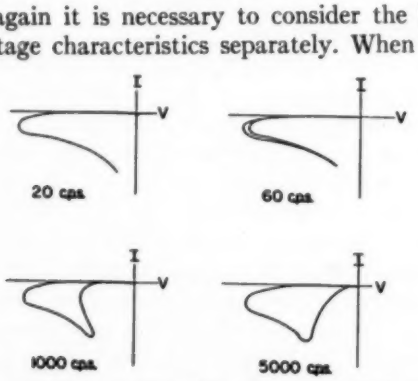


FIG. 15. Appearance of inverse characteristic on oscilloscope at various frequencies.

Measurements by R. N. Smith of the contact capacitance of high back voltage crystals gave values of the order of 0.5 $\mu\mu$f, which is similar to that for crystals which perform well at 3000 Mc.

Time effects in the spreading resistance are shown by the experiments of Yearian[17] at various frequencies and amplitudes which indicate that the drop in rectification efficiency of high voltage crystals is due to an increase in the spreading resistance with frequency. This may correlate with the work of Bray, Lark-Horovitz, and Smith,[16] who find strong voltage and time effects in the bulk conductivity of high voltage germanium samples.

The negative resistance characteristic is much more sluggish than that at low voltage. With the proper circuit, it is easily possible to generate negative resistance oscillations, but the oscillation usually ceases above a frequency of 10^5 per sec., indicating a relaxation time of the order of 10 μsec. If the high voltage characteristic is measured in a circuit containing insufficient series resistance, the discontinuous jump that occurs takes about 10^{-5} sec.

When alternating current is used to view the negative characteristic, considerable hysteresis may be observed, even at 60 c. This hysteresis occurs in spite of the fact that the circuit has sufficient series resistance to avoid discontinuities. At sufficiently low frequency (see Fig. 15) the hysteresis substantially disappears, while at high frequencies it becomes progressively worse. A large loop may be observed also at 60 c if a great current swing over the peak is used. These time lags seem excessively large in view of the facts that discontinuities

occur in 10^{-5} sec., and negative resistance oscillations as high as 10^5 per sec. may be produced.

In order to obtain more quantitative information, square waves were used to modulate the grid of a current-limiting pentode with the crystal in the plate circuit. Since the plate resistance of the tube was high compared to the crystal resistance, a square current pulse was obtained, as could be observed across a resistance in series with the crystal. The voltage pulse across the crystal itself was also observed. Figure 16 illustrates typical results. When the current is kept well below the peak of the characteristic, the voltage response is essentially flat-topped and increases with the current. Above a current value corresponding to the peak, however, the voltage drops with time, reaching a final value in about 10–20 μsec. This final value becomes smaller as the current is raised, while the initial value remains essentially fixed. This indicates that the instantaneous current-voltage characteristic does not have much negative resistance, the voltage merely reaching a limiting value as the current is raised. However, if 10^{-5} sec. is allowed for equilibrium, the negative resistance sets in, presumably due to heating at the contact.

In Part (6) of Fig. 16 the current pulse does not go to zero, but alternately jumps from the value B_1 (below the peak) to B_3 (above the peak). As the drop from B_3 to B_1 takes place, the voltage initially drops to a low value but largely recovers in the same time as required for the voltage to drop in going over the peak. However, the recovery is not complete in this time; even at 200 μsec. the voltage continues to rise. This slow effect, which is possibly due to limiting factors in the conduction of heat away from the contact, accounts for the large hysteresis observed in the current-voltage characteristic at audiofrequencies.

An interpretation of what is going on in Part (6) of Fig. 16 is illustrated in Fig. 17. Curve (1) represents the equilibrium (d.c.) characteristic. Curve (2) is the instantaneous characteristic at a crystal temperature corresponding to point C on the d.c. characteristic (current B_1). Curve (3) is the instantaneous characteristic corresponding to point E (current B_3). As the current alternates between B_1 and B_3 the voltage follows the path $CDEFC$, the transitions CD and EF being instantaneous while DE requires about 10 μsec. and FC is largely completed in 10 μsec., but requires much longer for full completion.

Some particular crystals have sharp breaks in their current-voltage characteristics which have much smaller relaxation times (about 1 μsec. or less). These effects are probably of a somewhat different nature from the ordinary negative resistance effect.

Boyarsky, Smith, and Yearian[25] made determinations of peak inverse voltage using megacycle frequencies

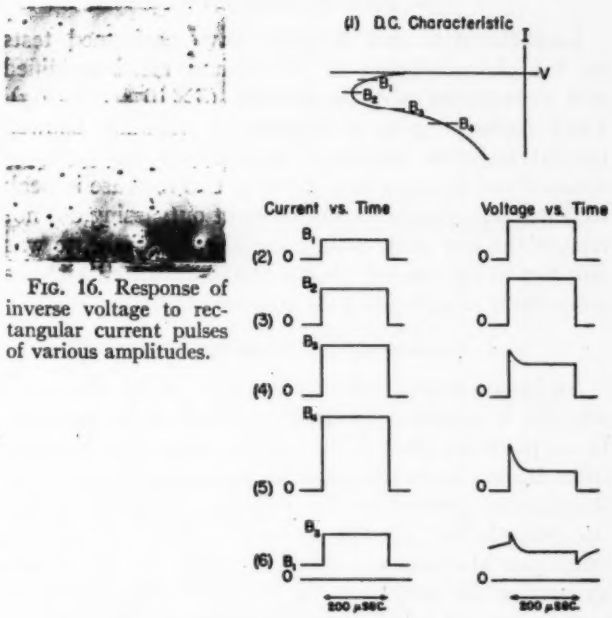


FIG. 16. Response of inverse voltage to rectangular current pulses of various amplitudes.

and measuring the maximum voltage which would be developed across the crystal. For crystals which were not power treated, the peak values were generally different from the d.c. peaks, by as much as a factor of two. Values lower or greater than the d.c. values could be obtained, depending upon whether or not current was allowed to pass in the forward direction during the measurement. The effect of power treatment was to greatly reduce the discrepancy between the peaks at radiofrequency and d.c.

5. *Performance Tests*

Extensive tests have been made to determine the performance of high voltage germanium rectifiers in various applications such as second detectors, d.c. restorers, and diode modulators. The effects of humidity, shock, pulses, long operation, temperature, and frequency on performance are given in great detail in the indicated reports.[26]

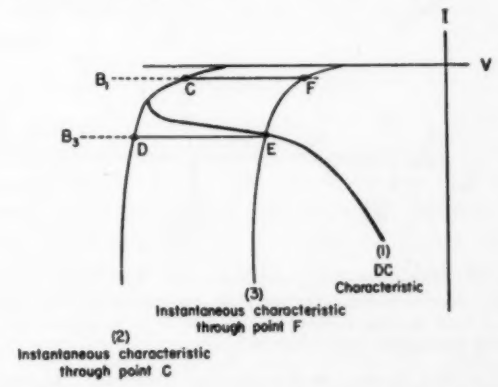


FIG. 17. Interpretation of sequence of events in Part 6 of Fig. 16.

[25] Boyarsky, Smith, and Yearian, "Properties of Germanium High Back Voltage Rectifier Units," NDRC Report 14-413, Purdue University (March 19, 1945).

[26] References 4, 14, 17, 19, 25. See also: R. N. Smith and H. J. Yearian, "Test Equipment for Germanium Second Detector

6. *Effect of Atmosphere*

Lark-Horovitz and Whaley[3] have performed tests on the characteristics of germanium melts solidified and maintained in high vacuum (3×10^{-6} mm Hg). Point contacts to such samples (in vacuum) showed normal negative resistance behavior in the back resistance and as much as a factor of two increase in peak back voltage. Reevacuation without outgassing did not restore the low peak voltage, so that the increase was not due to the cooling ability of the admitted air, but more likely to a firmly held adsorbed layer.

7. *Germanium-Germanium Contacts*

An interesting experiment is one in which the metal whisker is replaced by another piece of germanium. If one piece is a sharp point and the other a flat surface, there is usually rectification, as Benedicks[1] found. The sharp point, presumably due to extreme pressure at the tip, behaves like a metal in regard to the sign of rectification. Mechanical asymmetry may be minimized by cutting the samples in the form of wedges and contacting these at right angles, as in Fig. 18. This gives a small area of contact and allows the choosing of many points on each piece. The characteristics observed are shown in Fig. 19. Curve *A* is observed when one piece is contacted with a metal. (Actually, a wedge of germanium of very low resistivity is used in place of a metal, in order that the mechanical conditions be unchanged, since the different hardness of the metal would affect the results.) When two high voltage pieces are contacted, the result illustrated by curve *B* is obtained, showing a high voltage characteristic in both directions.

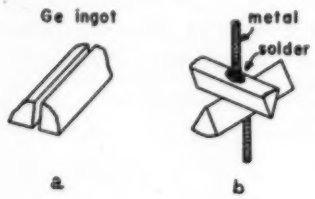


FIG. 18. Method of cutting and contacting samples for experiments on Ge-Ge contacts.

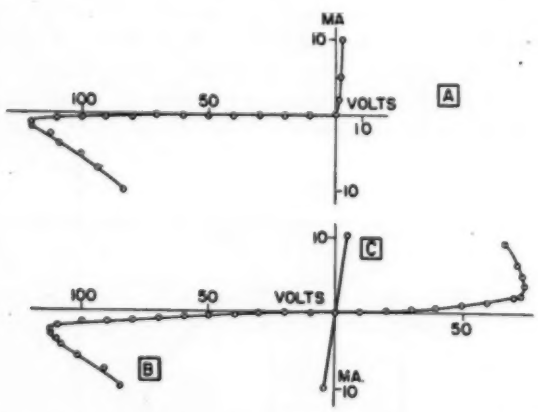


FIG. 19. Current-voltage characteristics for Ge-Ge contacts: (a) High inverse voltage Ge vs. very low resistance Ge. (b) Two pieces of high inverse voltage Ge. (c) Curve to be expected if only spreading resistance were observed.

If the two pieces of Ge come from samples having different peak voltages, an unsymmetrical characteristic is obtained, each crystal exhibiting its own peak. One may therefore set up a rectification series[27] where each crystal will rectify with a given polarity against all those below it in the series and with opposite polarity against those above.

According to the natural barrier layer theory, where the space charge barrier is due to the difference in work function of the metal and the semiconductor, such a contact should have no barrier and would be expected to have low resistance in both directions (equal to the sum of the spreading resistances). The fact that high voltage characteristics are obtained in both directions indicates that the barrier already exists in each crystal before contact. Both these experiments and the vacuum work of Lark-Horovitz and Whaley are in agreement with the ideas of Bardeen[28] on the role of surface states in rectification.

8. *Photoelectric Effects*

High inverse voltage germanium crystals in general exhibit photoelectric effects.† Illumination of the contact causes changes in current at a given applied voltage, as well as a photo-e.m.f. when there is no applied voltage. The peak inverse voltage also drops slightly.

The geometry of the point contact is such that only the part of the barrier which extends around the edge of the contact receives light which can be effective in producing a barrier layer photo-effect. Assuming that the thickness of the barrier layer is small compared with the contact radius, the sensitive area is a circular ring of width about equal to the barrier thickness. It might therefore be possible to obtain a measure of the barrier thickness by observation of the photoelectric effect.

Short circuit photo-currents were measured for crystals of a wide range of peak voltages, keeping constant contact force and illumination, with no voltage applied. Each crystal was freshly etched and for each point, the peak inverse voltage was measured first (no other power treatment was applied), then the low voltage d.c. characteristic (at 10 mv) to obtain the contact resistance. Short light pulses were then applied and the current pulse through a small resistance was amplified and observed on an oscilloscope. In the case of very low inverse voltage crystals, the photo-effect was so feeble that it was necessary to measure the open-circuit photovoltaic e.m.f. and compute the short circuit photo-current by dividing by the contact resistance. For the high inverse voltage crystals, this

Units," NDRC Report 14-394, Purdue University (January 25, 1945); and Boyarsky, Smith, and McDonald, "Dependence of Performance of Germanium Second Detector Units on Bias and Video Load," NDRC Report 14-416, Purdue University (March 28, 1945).

[27] W. H. Brattain, Phys. Rev. 69, 682 (1946).

[28] J. Bardeen, Phys. Rev. 71, 717 (1947).

† First observed by P. R. Bell at the M.I.T. Radiation Laboratory while testing some Purdue crystals.

procedure was not valid since the photovoltaic e.m.f. was several times e/kT and the nonlinearity of the contact resistance would have to be taken into account. However, for these crystals, it was easy to measure the photo-current in a small series resistor.

Figure 20 shows the correlation of photo-current with peak inverse voltage. This would seem to confirm the idea that the barrier thickness is larger for high voltage crystals. For barriers thin compared with the optical absorption depth, the photo-current should be proportional to the square of the thickness.

Photovoltaic cells of considerable sensitivity may be prepared by coating the surface of a high inverse voltage crystal with a thin (transparent) metallic film.

Further studies on photoelectric effects in germanium crystals have been made,[29] but these will be reported separately.

III. DISCUSSION

There are several features which distinguish germanium crystals, having high inverse voltage properties, from those of the low inverse voltage type. High peak voltage is apparently associated with poorer frequency response, larger excess noise, and nonlinear spreading resistance. There is also greater mechanical stability and an enhanced photoelectric sensivitity.

The photo-effect and mechanical stability are evidenced for thicker barriers in high inverse voltage units. In spite of the fact that a good deal of impurity is added, most of the impurity seems to be inert and good high voltage crystals generally have higher resistivity. The smaller values of N can be expected to give larger d values.

It would be desirable to obtain direct determinations of barrier thickness by means of capacity measurements. The method used by Smith[13] is not feasible, since R_s is not independent of voltage and frequency. However, by biasing in the reverse direction and measuring impedance *vs.* frequency, this should be possible.

On the basis of N values calculated from Hall effect determinations, the natural barrier thickness for a typical sample ($N=10^{15}$, $\epsilon=16$, $\phi=0.5$ v) at $V=0$ would be about 10^{-4} cm. At an applied negative voltage of 100 v, the barrier thickness would be ten times as thick ($d\sim[\phi+V]^{\frac{1}{2}}$) and the field strength would be of the order of 10^5 v/cm, an extremely high value. While a solid piece of germanium could probably not withstand such a field, the barrier is a limited number of mean free paths thick[4] and there is less opportunity for cascading effects to occur.

The mechanism of the inverse negative resistance characteristic seems intimately connected with heating at the contact. Well before the peak is reached, however, at current and voltage values corresponding to small amounts of power and negligible contact temperature rise, the current begins to rise rapidly with voltage.

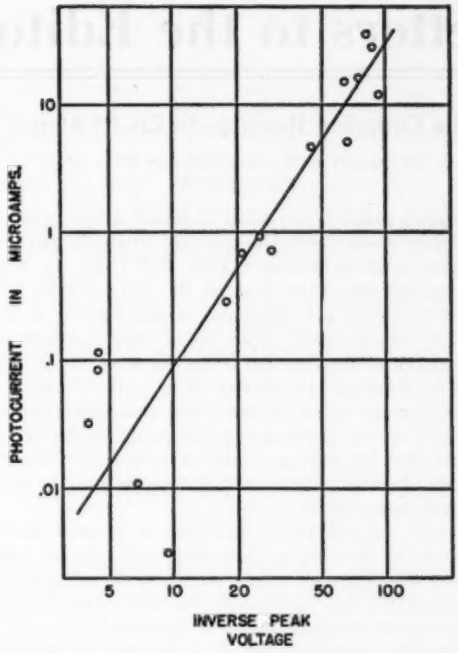


FIG. 20. Short circuit photovoltaic current at fixed light intensity *vs.* peak inverse voltage for various crystals.

An important fact which is apparent in the experiments on power treatment, temperature dependence, and contact force, is the independence of the reverse characteristic at low voltage from the peak characteristic. This might indicate that the rise in current as the peak is approached is not due to breakdown of the contact barrier itself. A possibility which might be considered seriously is that breakdown occurs around the edge of the contact due to surface conductivity.[30]

It is difficult at the present time to make any quantitative calculations of high voltage characteristics. Some attempts have been made,[31] but the explanation of the behavior in the forward direction is uncertain because of the complex nature of the spreading resistance, while the inverse characteristic may be solved only after further data on surface conductivity are obtained.

ACKNOWLEDGMENT

The author is greatly indebted to the members of the semiconductor group at Purdue, especially K. Lark-Horovitz, R. N. Smith, R. M. Whaley, and H. J. Yearian, for their unfailing stimulation and guidance.

This work was carried out initially under the research program of the OSRD and continued after the war under the auspices of the Signal Corps.

[29] S. Benzer, "Photoelectric Effects in Germanium," NDRC Report 14–580, Purdue University (October 31, 1945); "Excess-defect germanium contacts," Phys. Rev. 72, 1267 (1947).

[30] Evidence for surface conductivity in germanium is given by W. H. Brattain and J. Bardeen, Phys. Rev. 74, 231 (1948).

[31] (a) S. Benzer, "The High Voltage Germanium Rectifier II," NDRC Report 14–375, Purdue University (December 26, 1944), (b) H. C. Torrey, private communication. (c) B. Goodman. "Double-Valued Characteristics of Crystal Rectifiers," NDRC Report, University of Pennsylvania (1945).

Letters to the Editor

The Ordering Reaction in Co-Pt Alloys†

J. B. Newkirk,* A. H. Geisler,* and D. L. Martin**
March 2, 1949

An ordering reaction can occur in binary alloys of cobalt and platinum whose composition is near 50 atomic percent. The maximum temperature of order is about 825°C for the 50 atomic percent alloy and lower for those off this composition. No other reaction occurs below the maximum temperature of order. The unit cell is face-centered cubic above this temperature and ordered face-centered tetragonal below. In some respects this reaction has for its prototype the one found in the CuAu alloy.

Evidence is given which indicates that at certain temperatures and compositions the ordering reaction proceeds through a two-phase stage that by holding within a measurable temperature range discreet regions of order and of disorder may be caused to exist together in equilibrium.

On the basis of preliminary evidence, it appears that at an early stage of the ordering process, coherency between regions of order and of disorder may exist. Lattice straining, induced as a consequence of this, may account for the unusual physical properties which develop during the course of the ordering process. Thus, the process may resemble that of solid solution percipitation (aging) in its effect on certain physical properties.

Further study of the alloy is in progress.

† This letter is part of the Special Section on the Pittsburgh X-Ray and Electron Diffraction Conference which appears on pages 725–746 of this issue.

* Graduate Student, Department of Metallurgical Engineering, Carnegie Institute of Technology, Pittsburgh 13, Pennsylvania.
** Research Associate, Research Laboratory, General Electric Company, Schenectady, New York.

Impedance Synthesis without Use of Transformers

R. Bott and R. J. Duffin
Department of Mathematics, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
December 13, 1948

Let $Z(s)$ be termed a B(rune) function if: (1) it is a rational function; (2) it is real for real s; and (3) the real part of Z is positive when the real part of s is positive. In his significant thesis, O. Brune[1] shows that the driving-point impedance of a passive network is a B function of the complex frequency variable s. Conversely, he shows that any B function can be realized by some passive network and gives rules for constructing such a network. In this synthesis he is forced to employ transformers with perfect coupling. It is recognized by Brune and others that the introduction of perfect transformers is objectionable from an engineering point of view. Prior to Brune, R. M. Foster[2] had shown how to synthesize the driving-point impedance of networks containing no resistors by simple series-parallel combinations of inductors and capacitors. This note gives a similar synthesis of an arbitrary impedance by series-parallel combinations of inductors, resistors, and capacitors.

A B function can be expressed as the ratio of two polynomials without common factor. Let the "rank" be the sum of the degrees of these polynomials. Obviously any B function of rank O can be synthesized. Suppose, then, it has been shown that all B functions of rank lower than n can be synthesized, and let $Z(s)$ be a B function of rank n. Brune gives four rules for carrying out a mathematical induction to a B function of lower rank:

(a) If Z has a pole on the imaginary axis, then Z can be synthesized by a parallel resonant element in series with an impedance Z' of lower rank; $Z=1/(cs+1/ls)+Z'$ where l^{-1}, $c\geq 0$.

(b) If Z has a zero on the imaginary axis, then Z can be synthesized by a series resonant element in parallel with an impedance Z' of lower rank; $1/Z=1/(ls+1/cs)+1/Z'$ where l, $c^{-1}\geq 0$.

(c) If the real part of Z does not vanish on the imaginary axis, $Z=r+Z_0$ where r is a positive constant (to be interpreted as resistance) and Z_0 is a B function of no greater rank than Z.

Brune's fourth rule, (d), which employs the perfect transformer, we replace by the following procedure:

(d′) If none of these reductions are possible, there exists a $w>0$ such that $Z(iw)$ is purely imaginary. First assume that $Z(iw)=iwL$ with $L>0$. We now make use of a key theorem discovered by P. I. Richards.[3] Let k be a positive number, and let

$$R(s)=\frac{kZ(s)-sZ(k)}{kZ(k)-sZ(s)}. \tag{1}$$

Then $R(s)$ is a B function whose rank does not exceed the rank of $Z(s)$. Richards states this theorem for $k=1$; the above form is an obvious modification, because $Z(ks)$ is also a B function. Let k satisfy the equation $L=Z(k)/k$. This is clearly always possible, because the function on the right varies from ∞ to 0 as k varies from 0 to ∞. With this choice of k, clearly $R(iw)=0$. Solving (1) for Z gives

$$\begin{aligned} Z(s)&=(1/Z(k)R(s)+s/kZ(k))^{-1}+(k/Z(k)s+R(s)/Z(k))^{-1}\\ &=(1/Z_1(s)+Cs)^{-1}+(1/Ls+1/Z_2)^{-1}. \end{aligned} \tag{2}$$

Here $Z_1(s)=kLR(s)$, $Z_2(s)=kL/R(s)$, $C=1/k^2L$. Since Z_1 is a B function with a zero on the imaginary axis, it can be synthesized. Likewise, Z_2 is a B function with a pole on the imaginary axis and can be synthesized. $Z(s)$ is therefore synthesized by two networks in series. The first network consists of the impedance Z_1 in parallel with a capacitor C, and the second network consists of the impedance Z_2 in parallel with an inductor L. In the case that $Z(iw)=-iwL$, similar considerations applied to the function $1/Z$ show that Z is synthesized by two networks in parallel. The synthesized network finally resulting has the configuration of a tree whose branches are ladder networks.

Richards[4] has sought necessary and sufficient conditions for the driving-point impedance of resistor-transmission-line circuits by means of an ingenious transformation of the Brune theory. The perfect transformers, which are again found to be objectionable, may be dispensed with by the above procedure.

[1] O. Brune, J. Math. and Phys. 10, 191–236 (1931).
[2] R. M. Foster, Bell Syst. Tech. J. 3, 259 (1924).
[3] P. I. Richards, Duke Math. J. 14, 777–786 (1947).
[4] P. I. Richards, Proc. I.R.E. 36, 217–220 (1948).

An Improvement in the Shadow-Cast Replica Technique

S. J. Singer* and R. F. Petzold
*Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California***
May 6, 1949

Williams and Backus[1] have recently discussed in full detail the shadow-cast replica technique of electron microscopy. In the course of an investigation of the reactions of proteins in thin films,[2,3] we have performed some experiments with this technique embodying an improvement which we wish to report.

In this technique, a thin film of a metal such as chromium or uranium is deposited at an oblique angle onto the surface to be examined, by evaporation in a high vacuum. One method of removing this replica from the surface involves first, the deposition of a thin film (about 1000A) of parlodion on top of the metal film,

and second, the stripping of the parlodion and metal films together, with the aid of scotch tape.

Fine screens are placed on the scotch tape prior to stripping, so that afterwards, the screen with the replica film is disengaged and inserted in the microscope for observation.

The success of this stripping procedure depends on, among other things, the relative adhesions of the parlodion to the metal film, and the metal film to the underlying surface. We have found that ethyl cellulose*** films cast from a one percent solution in ethylene chloride are much more effective than parlodion films for this purpose, and moreover behave well in the electron microscope. In a number of instances, ethyl cellulose-backed chromium replicas were easily stripped from surfaces that completely retained those backed with parlodion. For example, three or five monolayers of barium stearate on a thin aluminum film on glass; one to six monolayers of bovine serum albumin on such a barium stearate surface; and a considerable number of similar surfaces, all behaved in this fashion. Electron micrographs of these replicas are shown elsewhere.[3] In no case that we encountered were parlodion-backed replicas stripped from surfaces that retained ethyl cellulose-backed films. Such surfaces included single monolayers of barium stearate applied to glass slides by the Blodgett-Langmuir technique;[4] ferric stearate films rubbed on glass;[4] and octadecylamine films deposited on glass from solution[5] in purified ligroin (b.p. 190–195 C).

In consideration of these results, we recommend that ethyl cellulose be used universally in shadow-cast replica techniques instead of parlodion. Since we do not have the opportunity to develop this technique further at present, we suggest that ethyl cellulose be tried particularly in those cases where parlodion-backed films are stripped only erratically, such as with platinum-palladium films or uranium films on glass.[1]

* U. S. Public Health Service Postdoctoral Fellow.
** Contribution No. 1285.
[1] R. C. Williams and R. C. Backus, J. App. Phys. **20**, 98 (1949).
[2] A. Rothen, J. Biol. Chem. **168**, 75 (1947).
[3] S. J. Singer, to be published.
*** Obtained from Hercules Powder Company, Wilmington, Delaware, and designated N-22.
[4] K. B. Blodgett and I. Langmuir, Phys. Rev. **51**, 964 (1937).
[5] Bigelow, Pickett, and Zisman, J. Coll. Sci. **1**, 513 (1946).

The Subcooling of Liquid Metals

D. TURNBULL
General Electric Research Laboratory, Schenectady, New York
May 10, 1949

VONNEGUT[1] has shown that an aggregate of very small droplets (1 to 10 microns) of liquid tin, separated by the oxide skin surrounding each particle, must be cooled 110 to 120°C below the thermodynamic liquid-solid transformation temperature before solidification proceeds rapidly.

We have made some observations on the rate of solidification of aggregates of mercury and gallium droplets. Collections of mercury droplets were prepared by dispersing mercury in solutions of sodium oleate in alcohol and in solutions of iodine in alcohol. In the latter, each mercury particle was coated with an iodide film. The particles averaged about 200 microns in diameter in the sodium oleate solution and about 50 microns in the iodine solution. The transformation was followed dilatometrically using alcohol as an indicator.

Starting about 20° above the melting point, T_m, samples were cooled about 10° an hour. After the transformation was completed the aggregate was slowly warmed through the melting temperature. Dilatometer readings were taken at close intervals in both the cooling and heating cycles. Mercury aggregates prepared by the first procedure did not show any measurable transformation until a temperature 33° below T_m had been reached. At this temperature the transformation was sluggish and probably confined to the larger particles. The rate of solidification became rapid at a temperature 39° below T_m. Aggregates held 22 degrees below did not transform after two hours.

Aggregates separated by the iodide film did not transform at a rapid rate until a temperature 46° below T_m had been attained. At 43° below T_m there was no appreciable transformation during a 1-hour holding period, except for 10 percent of the sample (probably the larger droplets) that transformed immediately.

Aggregates of gallium droplets were dispersed in alcoholic sodium oleate and the particles were about the same size as those of mercury in the same solution. The rate of solidification of gallium did not become appreciable until the gallium had been supercooled 70°C.

In Table I the maximum subcooling that has been obtained in bulk and in aggregates of small droplets are compared for tin, gallium, and mercury. It is interesting to note that the ratio T/T_m listed in the table, where T is the lowest absolute temperature to which the liquid has been subcooled, is roughly constant. Also shown in the table are liquid-solid interfacial energies σ calculated by the method of Fisher, Hollomon, and Turnbull.[2]

TABLE I.

Metal	Maximum subcooling °C: Bulk	Maximum subcooling °C: Aggregates of small droplets	T/T_m	σ
Tin	31*	110**	0.78	65
Mercury	14*	46	0.80	23
Gallium	55	70	0.75	55

* Danilov and Neumark, Physik Zeits. Sowjet, union **12**, 313 (1937).
** See reference 1.

Calculations based on the theory of homogeneous nucleation[2] indicate that the increased subcooling attained by dispersing bulk samples into droplets cannot be accounted for by the decreased nucleation probability occasioned by decreasing the droplet size. Rather there is evidence that nucleation in bulk liquid samples is generally "catalyzed" at the surface of heterogeneties. In pure liquids these "nucleation catalysts" may be limited in number so that they are segregated on a small portion of the droplets when the sample is dispersed. These promote transformation only on the droplets in which they are localized. Transformation of particles containing no catalysts may occur by homogeneous nucleation which proceeds at measurable speeds at much lower temperatures.

[1] Vonnegut, J. Colloid Science **3**, 563 (1948).
[2] Fisher, Hollomon, and Turnbull, Science **109**, 168 (1949).

Measurement of the Dielectric Constant and Loss of Solids and Liquids by a Cavity Perturbation Method

GEORGE BIRNBAUM
National Bureau of Standards, Washington, D. C.
AND
JACQUES FRANEAU*
Faculté Polytechnique de Mons, Mons, Belgium
May 12, 1949

A CAVITY resonator method for the measurement of the dielectric constant and loss of small samples was given by Sproull and Linder.[1] This method is characterized by convenient measurement techniques, simple calculations, and direct extension to the measurement of the shunt impedance of a cavity. The present work was undertaken to extend the usefulness fo this method by increasing its accuracy with a sensitive technique for measuring small frequency differences.

1. Review of the theory.—Bethe and Schwinger have developde a perturbation theory[2] which gives the change in resonance frequency (f) and loaded Q(Q) of a cavity due to some small change in the cavity. They consider two cavities, 1 and 2 which differ

slightly due to the presence of some material in cavity 2. If the material has a complex dielectric constant (relative to vacuum) given by $\epsilon'-i\epsilon''$, and permeability equal to one, then

$$\frac{f_1-f_2}{f_2}=(\epsilon'-1)\frac{\int_v \mathbf{E}_1\cdot\mathbf{E}_2 dv}{2\int_V E_1^2 dV}, \tag{1}$$

$$\frac{1}{Q_2}-\frac{1}{Q_1}=\epsilon''\frac{\int_v \mathbf{E}_1\cdot\mathbf{E}_2 dv}{\int_V E_1^2 dV}, \tag{2}$$

where v and V are the volumes of the dielectric and cavity, respectively, and $\mathbf{E}$ is the electric field intensity. These equations are useful if $\mathbf{E}_2$ can easily be obtained from a known unperturbed field $\mathbf{E}_1$. As an example, they discuss the case where the surface of the dielectric is everywhere parallel to the electric field. With the further restriction that the dielectric ends with the field lines on the walls of the cavity, $\mathbf{E}_2$ is put equal to $\mathbf{E}_1$. The relative error due to this and other approximations in obtaining Eqs. (1) and (2) is of the order $(f_1-f_2/f_2)+(1/Q_2)$.

2. Method of Measurement.—In the present arrangement, a small cylindrical sample is placed parallel to $\mathbf{E}$ by inserting it through holes in the wider walls of a rectangular cavity ($Q=5000$, $f=8900$ Mc), operating in the TE_{106} mode. The electric field is given by

$$E_y=A\sin\frac{\pi x}{a}\sin\frac{6\pi z}{b}, \tag{3}$$

where a and b are the width and length of the cavity, respectively. Assuming that the sample is located where E_y is a maximum, Eqs. (1) and (2) simplify to

$$(f_1-f_2/f_2)=2(\epsilon'-1)(v/V), \tag{4}$$

$$1/Q_2-1/Q_1=4\epsilon''(v/V). \tag{5}$$

The loss is characterized by $\tan\delta=\epsilon''/\epsilon'$.

The quantities Q and (f_2-f_1) are measured by a method which presents the cavity resonance curve and a pair of calibrated variable frequency markers in the form of intensified spots on an oscilloscope.[2] A block diagram of the apparatus is shown in Fig. 1, where the components enclosed by the dotted lines represent a cavity Q-meter.[4] With certain restrictions such as a high Q cavity loaded with matched transmission lines, and a square-law detector, the pattern traced out on the oscilloscope is of the resonance form

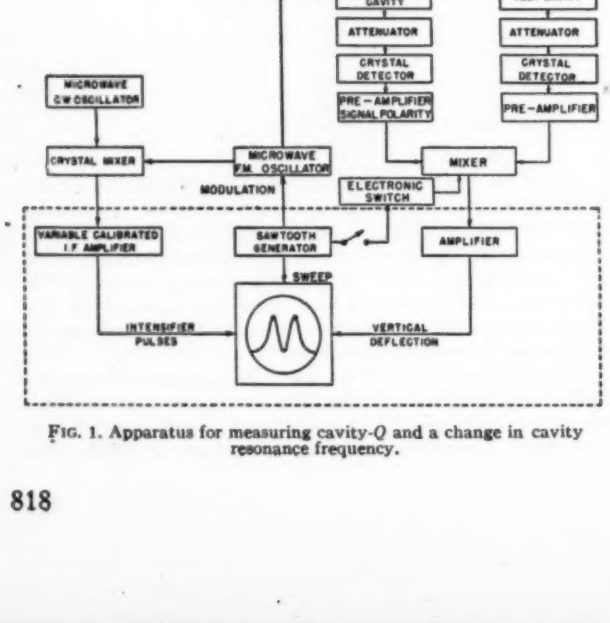


FIG. 1. Apparatus for measuring cavity-Q and a change in cavity resonance frequency.

TABLE I. Dielectric constant at 20°C.

Sample	Present method, 8900 Mc/s	Ref. 5, 9400 Mc/s
Polystyrene		
0.9-mm diam.	2.48	
2-mm diam.	2.48	
sample fitted with brass plugs	2.47	
*N-Heptane	1.93	1.9220
†Benzene	2.28	2.2835 (Analar)

TABLE II. Loss tangent.

Sample	Present method, 8900 Mc/s	Ref. 5, 9400 Mc/s
*N-Heptane	0.0003	0.00037
†Benzene	<0.0003	0.00050 (Analar)
Polystyrene	0.0012	

* National Bureau of Standards' standard sample 216, impurity 0.1 mole percent.
† Impurity less than 0.1 mole percent.

$[1+4Q_1^2(f-f_1)^2/f_1^2]^{-1}$. In the measurement of dielectric constant, the frequency difference between the reference and test cavities is obtained by aligning the frequency markers at the peaks of the two resonance curves. The tunable reference cavity is used to mark the frequency of the empty test cavity. For cavities with a Q of 5000, the standard deviation associated with a frequency difference measurement of approximately 1 to 15 Mc is ± 0.15 percent. The Q of the cavity is calculated as the mean of a set of measurements of the band width for several amplitudes, which are obtained with the frequency markers set at equal vertical displacements on each side of the resonance curve. The accuracy of the Q measurements rests upon the assumption that the oscilloscope pattern is of the resonance curve form. Hence the constancy of the aforementioned data serves to verify the correctness of the shape of the resonance curve. The resonance frequency of the test cavity is obtained from the calibration of the reference cavity by superimposing their resonance curves on the oscilloscope.[1] Liquid samples are contained in a thin-walled quartz tube, 2 mm in diameter.

3. Results.—Values obtained for ϵ' and $\tan\delta$ of some low loss liquids and solids are given in Tables I and II. An accuracy of ± 1 percent for ϵ' and $\pm 3\times 10^{-4}$ for $\tan\delta$ is assigned to these values. The accuracy is calculated from experimentally determined standard deviations of the parameters upon which ϵ' and $\tan\delta$ depend. The agreement between the present values and those obtained by another method which is considered to be more accurate,[5] may be taken as a check on the present method of measurement. Evidently no measurable error is introduced due to the fringing of the electric field at the holes in the cavity. This point was also checked by fitting a polystyrene sample with two brass plugs which accurately fit the holes through which the sample was inserted. In this case, the sample ends with the electric field lines on the walls of the cavity.

Improved accuracy for the $\tan\delta$ values may be expected by increasing the Q of the test cavity, and by increasing the volume of the sample, provided fringing at the holes and the inaccuracy in the theory do not become serious. If the sample diameter is increased, a correction becomes necessary in Eqs. (4) and (5) when the field is no longer accurately constant over the volume of the sample. Such a correction, amounting to 1 percent, was applied to ϵ' obtained for the 2-mm samples.

* Formerly a guest worker at the National Bureau of Standards, Washington, D. C.
[1] R. L. Sproull and E. G. Linder, Proc. I.R.E. **34**, 305 (1946).
[2] H. A. Bethe and J. Schwinger, NDRC Report D1-117, (March 1943). See also J. C. Slater, Rev. Mod. Physics **18**, 480–3 (1946).
[3] A similar technique for gases was reported by G. Birnbaum at a joint URSI-IRE Meeting, May 1947, Washington, D. C., and Lyons, Birnbaum, and Kryder, Phys. Rev. **74**, 1210(A), (1948).
[4] For circuit diagrams and a detailed description of the operation, see C. G. Montgomery, "Technique of Microwave Measurements" (McGraw-Hill Book Company, Inc., New York, 1947), Rad. Lab. Series, Vol. 11, sec. 6.29.
[5] Bleaney, Loubser, and Penrose, Proc. Phys. Soc. London **59**, 185 (1947).

Note on "The Mixing Properties of Non-Linear Condensers"

R. V. L. HARTLEY
Bell Telephone Laboratories, Murray Hill, New Jersey
June 9, 1949

IN a paper in the Journal[1] entitled "On the Mixing Properties of Non-Linear Condensers," A. Van der Ziel discusses the negative resistances which are associated with non-linear condensers and the resulting possibility of sustained oscillations at two frequencies, the sum of which is equal to that of the driving generator. He was obviously not familiar with my paper on the theory of these phenomena,[2] nor that by L. W. Hussey and L. R. Wrathal[3] in which the theoretical results are confirmed experimentally. These papers also give a solution in closed form for the oscillations at the various frequencies in terms of the circuit constants and the impressed voltage when the latter exceeds the threshold for oscillations. The circuit impedances are made such that the oscillations are negligible except at the frequencies considered. The writers referred to in the present paper seem also to have been unaware of the earlier papers.

[1] A. Van der Ziel, J. App. Phys. **19**, 999 (1948).
[2] R. V. L. Hartley, Bell Sys. Tech. J. **15**, 424 (1936).
[3] L. W. Hussey and L. R. Wrathal, Bell Sys. Tech. J. **15**, 441 (1936).

Note on "Qualitative Experimental Verification of the Change of Burning Rate of Rocket Powders with Radiation Path Length"

HENRI MURAOUR
L'ingenieur general des Poudres, Paris, France
April 18, 1949

I READ with deep interest the note under the title "Qualitative experimental verification of the change of burning rate of rocket powders with radiation path length."[1]

I was surprised by the results given, in Fig. 3, page 513. In fact, according to the study that I performed with G. Aunis,[2] the burning rate of a propellant, under constant pressure of 1000 kg/cm² might be calculated with the formula:

$$\log v = 1.37 + 0.27(T/1000),$$

v being the thickness lessening of the propellant in mm/seconds. If we take for rate of burning the rate of moving of the burning surface, normally to itself, these numbers must be divided by two.

On the other hand, we found[3] that the change in the rate of burning a propellant, as a function of the pressure, is linear and must be written: $v = a + bP$. Log$b$ is a linear function of T and a is close to 10 mm/sec. (v still being the thickness lessening in mm/sec., and P being expressed in kg/cm²).

According to the composition given by Penner,[4] we calculated an explosion temperature of about 4000°K (this calculation being done without taking the dissociation into account). From this temperature we obtained the following formula: $v = 10 + 0.272P$, or if we call r, as the author did, the normal rate of moving of the burning surface: $r = 5 + 0.136P$, r being expressed in mm/sec. and P in kg/cm².

Table I can be used to compare the calculated values with those Penner gave in Fig. 3, page 513.

This table shows the calculated values to be about twice as great as the values of Fig. 3. The value read on the graph would correspond to a powder with an explosion temperature close to 2600°K, therefore much lower than the previously calculated temperature of 4000°K.

TABLE I.

Pressure		Burning rate red on the scheme		Burning rate (4) calculated with our formula	
lb./in.²	kg/cm²	in./sec.	mm/sec.	in./sec.	mm/sec.
800	56	0.25	6.3	0.50	12.6
1200	84	0.30	7.6	0.65	16.4
1600	112	0.38	9.7	0.79	20.2
2400	168	0.50	12.7	1.09	27.8
3600	253	0.70	17.8	1.55	39.4

It must be noticed that the results published by Crawford and Alii[5] agree very satisfactorily with our formula (they used a powder of explosion temperature close to 4000°K and the method we previously indicated).[6]

We wonder as to the origin of these differences. Have we misunderstood Penner's paper or are the rates shown in Fig. 3 referring to a powder of explosion temperature close to 2600°K and not to the powder whose composition is indicated?[7]

[1] S. S. Penner, J. App. Phys. **19**, 511 (1948).
[2] H. Muraour and G. Aunis, Mémorial de l'Artillerie Française **20**, 637 (1946). H. Muraour, Chimie et Industrie **50**, 168 (1943).
[3] H. Muraour, Chimie et Industrie **47**, 602 (1942).
[4] S. S. Penner, J. App. Phys. **19**, 398 (1948).
[5] Crawford and Alii, Anal. Chem. **19**, 630 (1947).
[6] H. Muraour, Chimie et Industrie **47**, 476 (1942); **50**, 172 (1948).
[7] S. S. Penner, J. App. Phys. **19**, 392 (1948).

Reply to Muraour's Letter to the Editor

The results described in the three articles on the effect of radiation on the rate of burning of propellants[1] referred to double-base propellant grains. Because of security regulations, it was not possible to publish complete compositions for each of the powders listed. Since the powder composition is of secondary importance as regards the nature of the physical phenomena under discussion, it was felt that the description of a representative double-base powder composition would be sufficient to indicate to the reader the type of propellants with which the investigations were concerned. The burning rates shown in Fig. 3,[2] were obtained for a modified double-base propellant with an explosion temperature in the neighborhood of 2500°K. The calculations according to the empirical relation of Muraour and Aunis[3] between flame temperature and burning rate, described by Muraour, therefore indicate that the empirical equation correlates the flame temperature-burning rate relation for these particular modified double-base propellants as well as for other powders.

S. S. PENNER
Jet Propulsion Laboratory
California Institute of Technology

[1] S. S. Penner, J. App. Phys. **19**, 278, 392, 511 (1948).
[2] S. S. Penner, J. App. Phys. **19**, 513 (1948).
[3] H. Muraour and G. Aunis, Mémorial de l'Artillerie Française **20**, 637 (1946).

Erratum: Photo-Density Method of Determining the Relative Penetration of Diffused Sodium 24 Tracer into Glass

[J. App. Phys. **20**, 129 (1949)]
JAMES R. JOHNSON
Department of Ceramic Engineering, The Ohio State University, Columbus, Ohio
May 20, 1949

THE author has called attention to an error in Fig. 3. The values of the ordinate are reversed in their proper order. They should read, from top to bottom, 1, 10, 100 instead of 100, 10, 1.

New Books

Velocity-Modulated Thermionic Tubes

By A. H. W. Beck. Pp. 180+x, The Macmillan Company, New York, 1948. Price $3.75.

The purpose of this book is "to give a general introduction to velocity-modulation tubes and their mode of operation in such a form as to be readily intelligible to anyone with a reasonably adequate knowledge of pre-1939 radio technique." The book was not intended to be a formal textbook with full documentation, but rather a personal account of a division of research which is still progressing rapidly. The author writes as an expert of the wartime developments of his subject.

Gas Tables

By Joseph H. Keenan and Joseph Kaye. Pp. 238+x, John Wiley and Sons, Inc., New York, 1948. Price $5.00.

This book supersedes a previous one by the same authors entitled *Thermodynamic Properties of Air*. The properties of air have been re-examined and recalculated. The properties of combustion products of hydrocarbons and of the constituent gases have been added. Tables of functions useful in analyzing the flow of compressible fluids have been greatly extended.

For engineering application to the design of power apparatus, the precision of the present tables is of the same order as that of modern steam tables. The convenience is of a higher order.

The base temperature, for which enthalpy is zero, is not the same here as in the authors' previous work. It is taken to be zero on the absolute Fahrenheit scale in accordance with Rossini's data. Negative values of properties are thus avoided. The molal unit is used for all tables of thermodynamic properties except air.

Principles of Radar

By Denis Taylor and C. H. Westcott. Pp. 141+x, Plates V, The Macmillan Company, New York, 1948. Price $3.50.

This book is an exposition and survey of the principles underlying radar design, intended to help readers who have found the sudden release of so much detailed information perplexing, by indicating the common factors which underlie the many types of radar equipment developed and used during the war. The authors point out that methods used in radar date from as long ago as 1924, when the height of the ionosphere was first measured, and that at the outbreak of war in 1939 a few radar stations were already operating. Taylor is a superintendent of Telecommunications Research Establishment, British Ministry of Supply, and Westcott was at one time a senior scientific officer there.

The authors state in their preface: "The present volume is addressed to those engineers, physicists and mathematicians who have some knowledge of the fundamental principles and pre-war practice of radio and who wish to learn about radar. It should also be of service to advanced students of radio, as well as to many who have worked in some aspect of radar but who wish to see the subject presented as a whole. A knowledge of electromagnetic theory, the calculus, and the $j(=\nu-1)$ convention is assumed, as well as some acquaintance with normal radio practice."

Practical Analysis

By Fr. A. Willers. Pp. 422+x, Figs. 132, Dover Publications, New York, 1949. Price $6.00.

This is a translation by Robert T. Beyer of a book originally published in Germany in 1928 by an author whose work in the fields of the numerical, graphical, and in some measure of the instrumental, methods of practical analysis is widely recognized. Since numerical methods can generally be arranged so as to yield any desired accuracy, these methods are treated in great detail in this volume. On the other hand, Willers points out that graphical methods often give rapid results which are sufficiently accurate for the first approximations needed in most numerical procedures, and he presents a comprehensive study of these methods.

The chapters deal with the following general subjects: problems encountered in calculating with approximate numbers; standard methods of interpolation; methods of numerical differentiation and integration; practical determination of the roots of single algebraic equations and of systems of linear equations; empirical formulas and curve fitting; graphical and numerical integration of differential equations.

This volume is recommended as a convenient reference book.

Underwater Explosions

By Robert H. Cole. Pp. 437+ix, Princeton University Press, Princeton, New Jersey, 1948. Price $7.50.

The author states in his preface: "The content and purpose of this book are largely the result of research on underwater explosions carried out by many groups in the years 1941–46. Much of the present knowledge and understanding of this field was acquired because of the demands of these war years, and the few available discussions of the subject have become inadequate or obsolete. This book is an attempt to supply a reasonably comprehensive account which will be of use both to workers in the field of underwater explosions and to others interested in the basic physical processes involved."

The emphasis in this book has been placed on fundamental physical properties, and the most generally useful theoretical methods for discussion of shock-wave development and propagation and of later motion of the explosion products are developed from first principles. The author begins with a qualitative description of an underwater explosion and progresses to a general discussion of basic hydrodynamical relations. After a consideration of the detonation process as a physical event, he treats the theory of the shock wave, including discussion of the Kirkwood-Bethe Theory. He next describes various pressure gauges, the photography of underwater explosions, shock-wave measurements, motion of the gas sphere, and secondary pressure waves.

Quantum Mechanics

By Leonard I. Schiff. Pp. 404+xii, Figs. 30, McGraw-Hill Book Company, Inc., New York, 1948. Price $5.50.

Designed as a college textbook for senior mathematics and physics majors and graduate students, this book seeks to provide a comprehensive treatment of the subject of quantum mechanics in a single volume of moderate size. The purpose of the book is threefold: to explain the physical concepts of quantum mechanics, to describe the mathematical formalism, and to present illustrative examples of both the ideas and the methods. It is assumed that the reader is reasonably familiar with atomic structure, classical mechanics, and differential equations. In addition, he should have had some contact with electromagnetic theory and, for the latter part of the book, with the special theory of relativity.

Among the principal topics treated are:

The physical basis of quantum mechanics;
The Schrödinger wave equation and its application to energy level problems and collision theory;
The Heisenberg matrix formalism and transformation theory;
Approximation methods;
Effects of particle identity and spin;
Some applications to atomic, molecular and nuclear problems;
The relativistic wave equations of Schrödinger and Dirac;
The quantization of wave fields;
Electromagnetic radiation, treated both semi-classically and by means of quantum electrodynamics.

Problems are given at the end of each chapter.